(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 August 2001 (16.08.2001)

PCT

(10) International Publication Number
WO 01/59140 A1

(51) International Patent Classification[7]: C12N 15/82, 9/10

(21) International Application Number: PCT/GB01/00477

(22) International Filing Date: 8 February 2001 (08.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0002814.2 9 February 2000 (09.02.2000) GB

(71) Applicant *(for all designated States except US)*: **THE UNIVERSITY OF YORK** [GB/GB]; Heslington, York YO10 5DD (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BOWLES, Diana, Joy** [GB/GB]; The University of York, Heslington, York YO10 5DD (GB). **LI, Yi** [GB/GB]; The University of York, Heslington, York YO10 5DD (GB). **LIM, Eng-Kiat** [MY/GB]; The University of York, Heslington, York YO10 5DD (GB).

(74) Agent: **HARRISON GODDARD FOOTE**; Tower House, Merrion Way, Leeds LS2 8PA (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/59140 A1

(54) Title: TRANSGENIC CELLS EXPRESSING GLUCOSYLTRANSFERASE NUCLEIC ACIDS

(57) Abstract: The invention relates to transgenic cells which have been transformed with nucleic acids encoding glucosyltransferase polypeptides (GTases) and vectors for use in transformation of said cells.

Transgenic Cells Expressing Glucosyltransferase Nucleic Acids

The invention relates to transgenic cells which have been transformed with glucosyltransferase (GTases) nucleic acids.

GTases are enzymes which post-translationally transfer glucosyl residues from an activated nucleotide sugar to monomeric and polymeric acceptor molecules such as other sugars, proteins, lipids and other organic substrates. These glucosylated molecules take part in diverse metabolic pathways and processes. The transfer of a glucosyl moiety can alter the acceptor's bioactivity, solubility and transport properties within the cell and throughout the plant. One family of GTases in higher plants is defined by the presence of a C-terminal consensus sequence. The GTases of this family function in the cytosol of plant cells and catalyse the transfer of glucose to small molecular weight substrates, such as phenylpropanoid derivatives, coumarins, flavonoids, other secondary metabolites and molecules known to act as plant hormones. Available evidence indicates that GTases enzymes can be highly specific, such as the maize and *Arabidopsis* GTases that glucosylate indole-3-acetic acid (IAA).

The production and use of paper has increased in the last 10 years. For example, between 1989 and 1999 the production of paper and board in the UK has increased from 4.6 to 6.6 million tonnes. Worldwide consumption has also reflected a general increase in paper usage. For example, in the UK per capita consumption of paper is over 200kg per annum. In the USA this figure is over 300kg per annum.

Wood used in the paper industry is initially particulated, typically by chipping, before conversion to a pulp which can be utilised to produce paper. The pulping process involves the removal of lignin. Lignin is a major non-carbohydrate component of wood and comprises approximately one quarter of the raw material in wood pulp. The removal of lignin is desirable since the quality of the paper produced from the pulp is largely determined by the lignin content. Many methods have been developed to efficiently and cost effectively remove lignin from wood pulp. These methods can be chemical, mechanical or biological. For example, chemical methods to pulp wood are disclosed in WO9811294, EP0957198 and WO0047812. Although chemical methods are efficient means to remove lignin from pulp it is known that chemical treatments can result in degradation of polysaccharides and is expensive. Moreover, to remove residual lignin from pulp it is necessary to use strong bleaching agents which require removal before the pulp can be converted into paper. These agents are also damaging to the environment.

Biological methods to remove lignin are known. There are however disadvantages associated with such methods. For example it is important to provide micro-organisms (eg bacteria and/or fungi) which only secrete ligninolytic enzymes which do not affect cellulose fibres. This method is also very time consuming (can take 3-4 weeks) and expensive due to the need to provide bioreactors. Biological treatment can also include pre-treatment of wood chips to make them more susceptible to further biological or chemical pulping.

It is therefore desirable to provide further means by which lignin can be efficiently and cost effectively removed from wood pulp which do not have the disadvantages associated with prior art methods.

For the sake of clarity reference herein to transgenic means a plant which has been genetically modified to include a nucleic acid sequence not naturally found in said plant. For example, by over-expression of monolignol glucosyltransferases *in planta*, plant cell wall properties may be altered through increasing the flux through biosynthetic intermediates that are obligatory for incorporation and assembly of the lignin polymer. Conversely, reduction of the monolignol glucoside pools, such as through the use of nucleic acid comprising GTase sequences in antisense configuration may lead to altered properties through reducing the flux through specific intermediates. Changes in lignin composition, such as with decreased ratios of coniferyl alcohol to sinapyl alcohol are highly desirable in paper and pulping processes, because the more highly methylated lignin (sinapyl alcohol) is more easily removed during pulping processes (Chiang et al (1988) TAPPI J. 71, 173-176).

In some applications it may be desirable to change lignin composition and increase the lignin content of a plant cell to increase the mechanical strength of wood. This would have utility in, for example the construction industry or in furniture making.

Both lignin content and the level of cross-linking of polysaccharide polymers within plant cell walls, also play an important role in determining texture and quality of raw materials through altering the cell walls and tissue mechanical properties. For example, there is considerable interest in reducing cell separation in edible tissues since this would prevent over-softening and loss of juiciness. Phenolics, such as ferulic acid, play an important role in cell adhesion since they can be esterified to cell wall polysaccharides during synthesis and oxidatively cross-linked in the wall, thereby increasing rigidity. Most non-lignified tissues contain these phenolic components and their levels can be modified by altering flux through the same metabolic pathways as those culminating in lignin. Therefore, in the same way as for the

manipulation of lignin composition and content, GTase nucleic acid in sense and/or antisense configurations can be used to affect levels of ferulic acid and related phenylpropanoid derivatives that function in oxidative cross-linking. These changes in content have utility in the control of raw material quality of edible plant tissues.

Lignin and oxidative cross-linking in plant cell walls also play important roles in stress and defence responses of most plant species. For example, when non-woody tissues are challenged by pests or pathogen attack, or suffer abiotic stress such as through mechanical damage or UV radiation, the plant responds by localised and systemic alteration in cell wall and cytosolic properties, including changes in lignin content and composition and changes in cross-linking of other wall components. Therefore, it can also be anticipated that cell- or tissue-specific changes in these responses brought about by changed levels of the relevant GTase activities will have utility in protecting the plant to biotic attack and biotic/abiotic stresses.

GTases also have utility with respect to the modification of antioxidants. Reactive oxygen species are produced in all aerobic organisms during respiration and normally exist in a cell in balance with biochemical anti-oxidants. Environmental challenges, such as by pollutants, oxidants, toxicants, heavy metals and so on, can lead to excess reactive oxygen species which perturb the cellular redox balance, potentially leading to wide-ranging pathological conditions. In animals and humans, cardiovascular diseases, cancers, inflammatory and degenerative disorders are linked to events arising from oxidative damage.

Because of the current prevalence of these diseases, there is considerable interest in anti-oxidants, consumed in the diet or applied topically such as in UV-screens. Anti-oxidant micronutrients obtained from vegetables and fruits, teas, herbs and medicinal plants are thought to provide significant protection against health problems arising from oxidative stress. Well known anti-oxidants from plant tissues include for example: quercetin, luteolin, and the catechin, epicatechin and cyanidin groups of compounds.

Caffeic acid (3,4-dihydroxycinnamic acid) is a further example of an anti-oxidant with beneficial therapeutic properties.

Certain plant species, organs and tissues are known to have relatively high levels of one or more compounds with anti-oxidant activity. Greater accumulation of these compounds in those species, their wider distribution in crop plants and plant parts already used for food and

drink production, and the increased bioavailability of anti-oxidants (absorption, metabolic conversions and excretion rate) are three features considered to be highly desirable.

It will be apparent that changed levels of the relevant GTase activities capable of glucosylating anti-oxidant compounds *in planta* will allow the production of anti-oxidants with beneficial properties. GTase sequences can also be expressed in prokaryotes or simple eukaryotes, such as yeast, to produce enzymes for biotransformations in those cells, or as *in vitro* processing systems.

Statements of Invention

According to an aspect of the invention there is provided a transgenic cell comprising a nucleic acid molecule which encodes a polypeptide which has:

i) glucosyltransferase activity:

ii) is selected from the group comprising sequences of Figures 1A. 2A. 3A, 4A, 5A, 6A, 7A, 8A, 9A, 10A, 11A, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32

iii) nucleic acids which hybridise to the sequences represented in (ii) above; and

iv) nucleic acid sequences which are degenerate as a result of the genetic code to the sequences defined in (i) and (ii) above.

In a further preferred embodiment of the invention said nucleic acid molecule anneals under stringent hybridisation conditions to the sequence presented in Figures 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A, 10A, 11A, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29. 30, 31, 32

More preferably still said nucleic acid molecule is selected from Figures 7A, 8A, 9A, 10A, 15, 18, 19, 28 or 31.

Stringent hybridisation/washing conditions are well known in the art. For example, nucleic acid hybrids that are stable after washing in 0.1xSSC,0.1% SDS at 60°C. It is well known in the art that optimal hybridisation conditions can be calculated if the sequence of the nucleic acid is known. For example, hybridisation conditions can be determined by the GC content of the nucleic acid subject to hybridisation. Please see Sambrook et al (1989) Molecular Cloning; A Laboratory Approach. A common formula for calculating the stringency

conditions required to achieve hybridisation between nucleic acid molecules of a specified homology is:

$T_m = 81.5^0 C + 16.6 \text{ Log } [Na^+] + 0.41[\% G + C] - 0.63$ (%formamide).

In a preferred embodiment of the invention said transgenic cell is a eukaryotic cell. Preferably said eukaryotic cell is a plant cell or yeast cell.

In an alternative embodiment of the invention said transgenic cell is a prokaryotic cell.

In a further preferred embodiment of the invention the nucleic acid molecule is selected from the group comprising: antisense sequences of the sequences of any one of Figures 1C, 2C, 3C, 4C, 5C, 6C, 7C, 8C, 9C, 10C and 11C or parts thereof, or antisense sequences of the sense sequences presented in Figures 12-32. More preferably still said antisense sequence is selected from Figure 7C or 9C

In a further preferred embodiment of the invention said nucleic acid is cDNA.

In a yet further preferred embodiment of the invention said nucleic acid is genomic DNA.

In yet still a further preferred embodiment of the invention said plant is a woody plant selected from: poplar; eucalyptus; Douglas fir; pine; walnut; ash; birch; oak; teak; spruce. Preferably said woody plant is a plant used typically in the paper industry, for example poplar.

Methods to transform woody species of plant are well known in the art. For example the transformation of poplar is disclosed in US4795855 and WO9118094. The transformation of eucalyptus is disclosed in EP1050209 and WO9725434. Each of these patents is incorporated in their entirety by reference.

In a still further preferred embodiment of the invention said plant is selected from: corn (Zea mays), canola (Brassica napus, Brassica rapa ssp.), alfalfa (Medicago sativa), rice (Oryza sativa), rye (Secale cerale), sorghum (Sorghum bicolor, Sorghum vulgare), sunflower (helianthus annuas), wheat (Tritium aestivum), soybean (Glycine max), tobacco (Nicotiana tabacum), potato (Solanum tuberosum), peanuts (Arachis hypogaea), cotton (Gossypium hirsutum), sweet potato (Iopmoea batatus), cassava (Manihot esculenta), coffee (Cofea spp.),

coconut (Cocos nucifera), pineapple (Anana comosus), citris tree (Citrus spp.) cocoa (Theobroma cacao), tea (Camellia senensis), banana (Musa spp.), avacado (Persea americana), fig (Ficus casica), guava (Psidium guajava), mango (Mangifer indica), olive (Olea europaea), papaya (Carica papaya), cashew (Anacardium occidentale), macadamia (Macadamia intergrifolia), almond (Prunus amygdalus), sugar beets (Beta vulgaris), oats, barley, vegetables and ornamentals.

Preferably, plants of the present invention are crop plants (for example, cereals and pulses, maize, wheat, potatoes, tapioca, rice, sorghum, millet, cassava, barley, pea, and other root, tuber or seed crops. Important seed crops are oil-seed rape, sugar beet, maize, sunflower, soybean, and sorghum. Horticultural plants to which the present invention may be applied may include lettuce, endive, and vegetable brassicas including cabbage, broccoli, and cauliflower, and carnations and geraniums. The present invention may be applied in tobacco, cucurbits, carrot, strawberry, sunflower, tomato, pepper, chrysanthemum.

Grain plants that provide seeds of interest include oil-seed plants and leguminous plants. Seeds of interest include grain seeds, such as corn, wheat, barley, rice, sorghum, rye, etc. Oil-seed plants include cotton, soybean, safflower, sunflower, Brassica, maize, alfalfa, palm, coconut, etc. Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava been, lentils, chickpea, etc.

According to a further aspect of the invention there is provided a vector comprising the nucleic acid according to the invention operably linked to a promoter.

"Vector" includes, *inter alia*, any plasmid, cosmid, phage or *Agrobacterium* binary vector in double or single stranded linear or circular form which may or may not be self-transmissable or mobilizable, and which can transform a prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g. autonomous replicating plasmid with an origin of replication ie an episomal vector).

Suitable vectors can constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. For further details see, for example, Molecular Cloning: Laboratory Manual: 2nd edition, Sambrook et al. 1989, Cold Spring Habor

Laboratory Press or Current Protocols in Molecular Biology, Second Edition. Ausubel et al. Eds.. John Wiley & Sons, 1992.

Specifically included are shuttle vectors by which is meant a DNA vehicle capable, naturally or by design, of replication in two different host organisms, which may be selected from actinomycetes and related species, bacteria and eukaryotic (e.g. higher plant, mammalian, yeast or fungal cells).

A vector including nucleic acid according to the invention need not include a promoter or other regulatory sequence, particularly if the vector is to be used to introduce the nucleic acid into cells for recombination into the gene.

Preferably the nucleic acid in the vector is under the control of, and operably linked to, an appropriate promoter or other regulatory elements for transcription in a host cell such as a microbial, (e.g. bacterial), or plant cell. The vector may be a bi-functional expression vector which functions in multiple hosts. In the case of GTase genomic DNA this may contain its own promoter or other regulatory elements and in the case of cDNA this may be under the control of an appropriate promoter or other regulatory elements for expression in the host cell.

By "promoter" is meant a nucleotide sequence upstream from the transcriptional initiation site and which contains all the regulatory regions required for transcription. Suitable promoters include constitutive, tissue-specific, inducible, developmental or other promoters for expression in plant cells comprised in plants depending on design. Such promoters include viral, fungal, bacterial, animal and plant-derived promoters capable of functioning in plant cells.

Constitutive promoters include, for example CaMV 35S promoter (Odell et al. (1985) Nature 313, 9810-812); rice actin (McElroy et al. (1990) Plant Cell 2: 163-171); ubiquitin (Christian et al. (1989) Plant Mol. Biol. 18 (675-689); pEMU (Last et al. (1991) Theor Appl. Genet. 81: 581-588); MAS (Velten et al. (1984) EMBO J. 3. 2723-2730); ALS promoter (U.S. Application Seriel No. 08/409,297), and the like. Other constitutive promoters include those in U.S. Patent Nos. 5,608.149; 5,608.144; 5,604.121; 5.569,597; 5,466,785; 5.399,680, 5,268,463; and 5,608,142.

Chemical-regulated promoters can be used to modulate the expression of a gene in a plant through the application of an exogenous chemical regulator. Depending upon the objective, the promoter may be a chemical-inducible promoter, where application of the chemical induced gene expression, or a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters are known in the art and include, but are not limited to, the maize In2-2 promoter, which is activated by benzenesulfonamide herbicide safeners, the maize GST promoter, which is activated by hydrophobic electrophilic compounds that are used as pre-emergent herbicides, and the tobacco PR-1a promoter, which is activated by salicylic acid. Other chemical-regulated promoters of interest include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter in Schena et al. (1991) Proc. Natl. Acad. Sci. USA 88: 10421-10425 and McNellis et al. (1998) Plant J. 14(2): 247-257) and tetracycline-inducible and tetracycline-repressible promoters (see, for example, Gatz et al. (1991) Mol. Gen. Genet. 227: 229-237, and US Patent Nos. 5,814,618 and 5,789,156, herein incorporated by reference.

Where enhanced expression in particular tissues is desired, tissue-specific promoters can be utilised. Tissue-specific promoters include those described by Yamamoto et al. (1997) Plant J. 12(2): 255-265; Kawamata et al. (1997) Plant Cell Physiol. 38(7): 792-803; Hansen et al. (1997) Mol. Gen. Genet. 254(3): 337-343; Russell et al. (1997) Transgenic Res. 6(2): 157-168; Rinehart et al. (1996) Plant Physiol. 112(3): 1331-1341; Van Camp et al. (1996) Plant Physiol. 112(2): 525-535; Canevascni et al. (1996) Plant Physiol. 112(2): 513-524; Yamamoto et al. (1994) Plant Cell Physiol. 35(5): 773-778; Lam (1994) Results Probl. Cell Differ. 20: 181-196; Orozco et al. (1993) Plant Mol. Biol. 23(6): 1129-1138; Mutsuoka et al. (1993) Proc. Natl. Acad. Sci. USA 90 (20): 9586-9590: and Guevara-Garcia et al (1993) Plant J. 4(3): 495-50.

"Operably linked" means joined as part of the same nucleic acid molecule, suitably positioned and oriented for transcription to be initiated from the promoter. DNA operably linked to a promoter is "under transcriptional initiation regulation" of the promoter. In a preferred aspect, the promoter is an inducible promoter or a developmentally regulated promoter.

Particular of interest in the present context are nucleic acid constructs which operate as plant vectors. Specific procedures and vectors previously used with wide success upon plants are described by Guerineau and Mullineaux (1993) (Plant transformation and expression vectors.

In: Plant Molecular Biology Labfax (Croy RRD ed) Oxford, BIOS Scientific Publishers, pp 121-148. Suitable vectors may include plant viral-derived vectors (see e.g. EP-A-194809). If desired, selectable genetic markers may be included in the construct, such as those that confer selectable phenotypes such as resistance to antibodies or herbicides (e.g. kanamycin, hygromycin, phosphinotricin, chlorsulfuron, methotrexate, gentamycin, spectinomycin, imidazolinones and glyphosate).

According to a further aspect of the invention there is provided a method of enhancing monolignol glucoside synthesis in a plant comprising causing or allowing expression of at least one GTase nucleic acid according to the invention in a plant. Preferably the plant is a woody plant species.

According to a further aspect of the invention there is provided a method of inhibiting monolignol glucoside synthesis in a plant comprising causing or allowing expression of at least one GTase antisense nucleic acid according to the invention in a plant. Preferably the plant is a woody plant species.

Inhibition of GTase expression may, for instance, be achieved using anti-sense technology.

In using anti-sense genes or partial gene sequences to down-regulate gene expression, a nucleotide sequence is placed under the control of a promoter in a "reverse orientation" such that transcription yields RNA which is complementary to normal mRNA transcribed from the "sense" strand of the target gene. See, for example, Rothstein *et al*, 1987; Smith *et al*, (1998), Nature 334, 724-726; Zhang et al (1992) The Plant Cell 4, 1575-1588, English *et al*. (1996) The Plant Cell 8, 179 188. Antisense technology is also reviewed in Bourque (1995), Plant Science 105, 125-149, and Flavell (1994) PNAS USA 91, 3490-3496.

According to a further aspect of the invention there is provided a nucleotide sequence encoding an antisense RNA molecule complementary to a sense mRNA molecule encoding for a polypeptide having a glucosyl transferase activity in the biosynthesis of at least a monolignol glucoside in lignin biosynthesis in a plant, which nucleotide sequence is under transcriptional control of a promoter and a terminator, both promoter and terminator capable of functioning in plant cells.

Suitable promoters and terminators are referred to hereinabove.

According to a further aspect of the invention there is provided a nucleotide sequence according to the invention comprising a transcriptional regulatory sequence, a sequence under the transcriptional control thereof which encodes an RNA which consists of a plurality of subsequences, characterised in that the RNA subsequences are antisense RNAs to mRNAs of proteins having a GTase activity in the lignin biosynthesis pathway in plant cells.

In particular, the said RNA subsequences are antisense RNAs to mRNAs of GTase having a GTase activity in the lignin biosynthesis pathway in plant cells, such as the GTase of Figs. 1-11 (C)

The nucleotide sequence may encode an RNA having any number of subsequences. Preferably, the number of subsequences lies between 2 and 7 (inclusive) and more preferably lies between 2-4 .

According to a further aspect of the invention there is provided a host cell transformed with nucleic acid or a vector according to the invention, preferably a plant or a microbial cell. The microbial cell may be prokaryotic (eg *Escherchia coli, Bacillus subtilis*) or eukaryotic (eg *Saccharomyces cerevisiae*).

In the transgenic plant cell the transgene may be on an extra-genomic vector or incorporated, preferably stably, into the genome. There may be more than one heterologous nucleotide sequence per haploid genome.

According to a yet further aspect of the invention there is provided a method of transforming a plant cell comprising introduction of a vector into a plant cell and causing or allowing recombination between the vector and the plant cell genome to introduce a nucleic acid according to the invention into the genome.

Plants transformed with a DNA construct of the invention may be produced by standard techniques known in the art for the genetic manipulation of plants. DNA can be introduced into plant cells using any suitable technology, such as a disarmed Ti-plasmid vector carried by Agrobacterium exploiting its natural gene transferability (EP-A-270355, EP-A-0116718, NAR 12(22):8711-87215 (1984), Townsend et al., US Patent No. 5,563,055); particle or microprojectile bombardment (US Patent No. 5,100,792, EP-A-444882, EP-A-434616; Sanford et al, US Patent No. 4,945,050; Tomes et al. (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment", in Plant Cell, Tissue and Organ Culture:

Fundamental Methods, ed. Gamborg and Phillips (Springer-Verlag, Berlin); and McCabe et al. (1988) Biotechnology 6: 923-926); microinjection (WO 92/09696, WO 94/00583, EP 331083, EP 175966, Green et al. 91987) Plant Tissue and Cell Culture, Academic Press, Crossway et al. (1986) Biotechniques 4:320-334); electroporation (EP 290395, WO 8706614, Riggs et al. (1986) Proc. Natl. Acad. Sci. USA 83:5602-5606; D'Halluin et al. 91992). Plant Cell 4:1495-1505) other forms of direct DNA uptake (DE 4005152, WO 9012096. US Patent No. 4,684,611, Paszkowski et al. (1984) EMBO J. 3:2717-2722); liposome-mediated DNA uptake (e.g. Freeman et al (1984) Plant Cell Physiol, 29:1353); or the vortexing method (e.g. Kindle (1990) Proc. Nat. Acad. Sci. USA 87:1228). Physical methods for the transformation of plant cells are reviewed in Oard (1991) Biotech. Adv. 9:1-11. See generally, Weissinger et al. (1988) Ann. Rev. Genet. 22:421-477; Sanford et al. (1987) Particulate Sciences and Technology 5:27-37; Christou et al. (1988) Plant Physiol. 87:671-674; McCabe et al. (1988) Bio/Technology 6:923-926; Finer and McMullen (1991) In Vitro Cell Dev. Biol. 27P:175-182; Singh et al. (1988) Theor. Appl. Genet. 96:319-324; Datta et al. (1990) Biotechnology 8:736-740; Klein et al. (1988) Proc. Natl. Acad. Sci. USA 85: 4305-4309; Klein et al. (1988) Biotechnology 6:559-563; Tomes, US Patent No. 5,240,855; Buising et al. US Patent Nos. 5,322, 783 and 5,324,646; Klein et al. (1988) Plant Physiol 91: 440-444; Fromm et al (1990) Biotechnology 8:833-839; Hooykaas-Von Slogteren et al. 91984). Nature (London) 311:763-764; Bytebier et al. (1987) Proc. Natl. Acad. Sci. USA 84:5345-5349; De Wet et al. (1985) in The Experimental Manipuation of Ovule Tissues ed. Chapman et al. (Longman, New York), pp. 197-209; Kaeppler et al. (1990) Plant Cell Reports 9:415-418 and Kaeppler et al. (1992) Theor. Appl. Genet. 84:560-566; Li et al. (1993) Plant Cell Reports 12: 250-255 and Christou and Ford (1995) Annals of Botany 75: 407-413;Osjoda et al. (1996) Nature Biotechnology 14:745-750, all of which are herein incorporated by reference.

Agrobacterium transformation is widely used by those skilled in the art to transform dicotyledonous species. Recently, there has been substantial progress towards the routine production of stable, fertile transgenic plants in almost all economically relevant monocot plants (Toriyama et al. (1988) Bio/Technology 6: 1072-1074; Zhang et al. (1988) Plant Cell rep. 7379-384; Zhang et al. (1988) Theor. Appl. Genet. 76:835-840; Shimamoto et al. (1989) Nature 338:274-276; Datta et al. (1990) Bio/Technology 8: 736-740; Christou et al. (1991) Bio/Technology 9:957-962; Peng et al (1991) International Rice Research Institute, Manila, Philippines, pp.563-574; Cao et al. (1992) Plant Cell Rep. 11: 585-591; Li et al. (1993) Plant Cell Rep. 12: 250-255; Rathore et al. (1993) Plant Mol. Biol. 21:871-884; Fromm et al (1990) Bio/Technology 8:833-839; Gordon Kamm et al. (1990) Plant Cell 2:603-618; D'Halluin et al. (1992) Plant Cell 4:1495-1505; Walters et al. (1992) Plant Mol. Biol. 18:189-200; Koziel et

al. (1993). Biotechnology 11194-200; Vasil, I.K. (1994) Plant Mol. Biol. 25:925-937; Weeks et al (1993) Plant Physiol. 102:1077-1084; Somers et al. (1992) Bio/Technology 10:1589-1594; WO 92/14828. In particular, Agrobacterium mediated transformation is now emerging also as an highly efficient transformation method in monocots. (Hiei, et al. (1994) The Plant Journal 6:271-282). See also, Shimamoto, K. (1994) Current Opinion in Biotechnology 5:158-162; Vasil, et al. (1992) Bio/Technology 10:667-674; Vain, et al. (1995) Biotechnology Advances 13(4):653-671; Vasil, et al. (1996) Nature Biotechnology 14: 702).

Microprojectile bombardment, electroporation and direct DNA uptake are preferred where Agrobacterium is inefficient or ineffective. Alternatively, a combination of different techniques may be employed to enhance the efficiency of the transformation process, e.g. bombardment with Agrobacterium-coated microparticles (EP-A-486234) or microprojectile bombardment to induce wounding followed by co-cultivation with Agrobacterium (EP-A-486233).

Plants which include a plant cell according to the invention are also provided.

In addition to the regenerated plant, the present invention embraces all of the following: a clone of such a plant, seed, selfed of hybrid progeny and descendants (e.g. F1 and F2 descendants).

According to a further aspect of the invention there is provided an isolated nucleic acid molecule obtainable from *Arabidopsis thaliana* which comprises a nucleic acid sequence encoding a polypeptide having

(1) GTase functionality; and

(2) is capable of adding a glucosyl group via an O-glucosidic linkage to form

(a) a glucosyl ester of at least one of :

cinnamic acid; *p*-coumaric acid; caffeic acid; ferulic acid; and sinapic acid; and/or

(b) a 4-O-glucoside of at least one of:

cinnamic acid; *p*-coumaric acid; caffeic acid; ferulic acid; sinapic acid; *p*-coumaryl aldehyde; coniferyl aldehyde; sinapyl aldehyde; *p*-coumaryl alcohol; coniferyl alcohol; and sinapyl alcohol.

In a further aspect of the invention there is provided a polypeptide encoded by an isolated nucleic acid molecule of the present invention wherein the said polypeptide is selected from the polypeptides of Figures 1B, 2B, 3B, 4B, 5B, 6B, 7B, 8B, 9B, 10B and 11B or functional variants and/or parts thereof. Preferably the polypeptide is selected from the group of polypeptides of Figures 2B, 3B, 4B, 6B, 7B and 9B or functional variants and/or parts thereof. Preferably still the polypeptide is selected from the group of polypeptides selected from Figures 2B, 3B, 7B and 9B or functional variants and/or parts thereof. Most preferably the polypeptide is one of the polypeptides shown in Figures 2B, 3B, 7B or 9B. Polypeptides encoded by the sense nucleic acid sequences presented in Figures 12 – 32 are also provided and readily derived from these sense sequences.

Variants of sequences having substantial identity or homology with the GTase molecules of the invention may be utilized in the practices of the invention. That is, the GTase of Figures 1A-11A may be modified yet still remain functional. Generally, the GTase will comprise at least about 40%-60%, preferably about 60%-80%, more preferably about 80%-95% sequence identity with a GTase nucleotide sequence of Figures 1A- 32 herein.

The activity of functional variant polypeptides may be assessed by transformation into a host capable of expressing the nucleic acid of the invention. Methodology for such transformation is described in more detail below.

In a further aspect of the invention there is disclosed a method of producing a derivative nucleic acid comprising the step of modifying any of the sequences disclosed above, particularly the coding sequence of Figures 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A, 10A, 11A, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32

Alternatively, changes to a sequence may produce a derivative by way of one or more of addition, insertion, deletion or substitution of one or more nucleotides in the nucleic acid, leading to the addition, insertion, deletion or substitution or one or more amino acids in the encoded polypeptide.

Other desirable mutations may be random or site directed mutagenesis in order to alter the activity (e.g. specificity) or stability of the encoded polypeptide or to produce dominant negative variants which may alter the flux through lignin biosynthetic pathways to alter the amount of lignin or an intermediate in the lignin biosynthetic pathway.

The invention will now be described with reference to the following Figures and Examples which are not to be construed as limiting the invention.

Scheme 1: The major intermediates in lignin biosynthesis pathway.

Figure 1A: Sense nucleotide sequence of A062. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 1B: The amino acid sequence of A062;

Figure 1C: The antisense nucleotide sequence of A062;

Figure 2A Sense nucleotide sequence of A320. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 2B The amino acid sequence of A320;

Figure 2C: The antisense nucleotide sequence of A320;

Figure 3A: Sense nucleotide sequence of A41. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 3B: The amino acid sequence of A41;

Figure 3C: The antisense nucleotide sequence of A41;

Figure 4A: Sense nucleotide sequence of A42. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 4B: The amino acid sequence of A42;

Figure 4C: The antisense nucleotide sequence of A42;

Figure 5A: Sense nucleotide sequence of A43. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 5B: The amino acid sequence of A43;

Figure 5C: The antisense nucleotide sequence of A43;

Figure 6A: Sense nucleotide sequence of A911. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 6B: The amino acid sequence of A911;

Figure 6C: The antisense nucleotide sequence of A911;

Figure 7A: Sense nucleotide sequence of A119. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 7B: The amino acid sequence of A119;

Figure 7C: The antisense nucleotide sequence of A119;

Figure 8A: Sense nucleotide sequence of A233. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 8B: The amino acid sequence of A233;

Figure 8C: The antisense nucleotide sequence of A233;

Figure 9A: Sense nucleotide sequence of A407. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 9B: The amino acid sequence of A407;

Figure 9C: The antisense nucleotide sequence of A407;

Figure 10A: Sense nucleotide sequence of A961. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 10B: The amino acid sequence of A961;

Figure 10C: The antisense nucleotide sequence of A961;

Figure 11A: Sense nucleotide sequence of A962. The coding region starts from the first nucleotide and ends at the last nucleotide;

Figure 11B: The amino acid sequence of A962;

Figure 11C: The antisense nucleotide sequence of A962.;

Figure 12: The sense nucleotide sequence of UGT71B5;

Figure 13 The sense nucleotide sequence of UGT71C3;

Figure 14 The sense nucleotide sequence of UGT71C5;

Figure 15 The sense nucleotide sequence of UGT71D1;

Figure 16 The sense nucleotide sequence of UGT73B1;

Figure 17 The sense nucleotide sequence of UGT73B2;

Figure 18 The sense nucleotide sequence of UGT73B4;

Figure 19 The sense nucleotide sequence of UGT73B5;

Figure 20 The sense nucleotide sequence of UGT73C1;

Figure 21 The sense nucleotide sequence of UGT731C;

Figure 22 The sense nucleotide sequence of UGT73C5;

Figure 23 The sense nucleotide sequence of UGT73C6;

Figure 24 The sense nucleotide sequence of UGT73C7;

Figure 25 The sense nucleotide sequence of UGT74F2;

Figure 26 The sense nucleotide sequence of UGT76E1;

Figure 27 The sense nucleotide sequence of UGT76E11;

Figure 28 The sense nucleotide sequence of UGT76E12;

Figure 29 The sense nucleotide sequence of UGT76E2;

Figure 30 The sense nucleotide sequence of UGT78D1;

Figure 31 The sense nucleotide sequence of UGT89B1;

Figure 32 The sense nucleotide sequence of UGT72B3;

Figure 33 shows recombinant GST-UGT71C1 fusion protein purified from *E. coli* using glutathione-coupled Sepharose. The protein (5 μg) was analyzed using 10% SDS-PAGE and was visualized with Coomassive staining;

Figure 34 shows three different glucose conjugates of caffeic acid, (caffeoyl-3-*O*-glucoside, caffeoyl-4-*O*-glucoside and 1-*O*-caffeoylglucose), obtained from the glucosyltransferase reactions containing the recombinant UGT71C1, UGT73B3 and UGT84A1 respectively. Each assay contained 1-2 μg of recombinant UGT, 1 mM caffeic acid, 5 mM UDP-glucose, 1.4 mM 2-mercaptoethanol and 50 mM TRIS-HCl, pH 7.0. The mix was incubated at 30 °C for 30 min and was analyzed by reverse-phase HPLC subsequently. Linear gradient (10-16%) of acetonitrile in H_2O at 1 ml/min over 20 min was used to separate the glucose conjugates from caffeic acid.

Figure 35A shows the pH optima of UGT71C1 glucosyltransferase activity measured over the range pH 5.5-8.0 in the reactions containing 50 mM buffer, 1 μg of UGT71C1, 1 mM caffeic acid, 5 mM UDP-glucose and 1.4 mM 2-mercaptoethanol. The mix was incubated at 30 °C for 30 min. The reaction was stopped by the addition of 20 μl of trichloroacetic acid (240 mg/ml) and was analyzed by reverse-phase HPLC subsequently. The specific enzyme activity was expressed as nanomoles of caffeic acid glucosylated per second (nkat) by 1 mg of protein in 30 min of reaction time at 30 °C. Figure 35B, the time course of UGT71C1 glucosyltransferase activity was studied by measuring the amount of caffeic acid glucosylated by 1 μg of UGT71C1 in 50 mM TRIS-HCl, pH 7.0. The reactions were carried out and analyzed as described in *A*;

Figure 36 shows UGT71C1 transgenic *Arabidopsis thaliana* plants and their ability to glucosylate caffeic acid; and

Figure 37 summarises the GTase activities of various GTase polypeptides with respect to various anti-oxidant substrates.

EXAMPLES

MATERIALS AND METHODS

Transformation of Woody Plant Species

The transformation of woody plant species is known in the art. See US4795855 and WO9118094; EP1050209 and WO9725434. Each of these patents are incorporated in their entirety by reference.

Transformation of Non-Woody Plant Species

Methods used in the transformation of plant species other than woody species are well known in the art and are extensively referenced herein.

Identification of GTase sequences

The GTase sequence identification was carried out using GCG software (Wisconsin package, version 8.1). Blasta programme was used to search *Arabidopsis* protein sequences containing a PSPG (plant secondary product UDP-glucose glucosyltransferase) signature motif (Hughes and Hughes (1994) DNA Sequence 5, 41-49) in EMBL and GenBank sequence database. The database information on the GTases described in the present invention are listed in Table 1.

Amplification and cloning of the GTase sequences

The GTase sequences were amplified from *Arabidopsis thaliana* Columbia genomic DNA with specific primers (Table 2), following standard methodologies (Sambrook et al (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). 50ng of genomic DNA isolated from Arabidopsis thaliana Columbia were incubated with 1 x pfu PCR buffer (Stratagene), 250 µM dNTPS, 50 pmole primer for each end, and 5 units of pfu DNA polymerase (Stratagene) in a total of 100 µl. The PCR reactions were carried out as outlined in the programme described in Table 3.

After PCR amplification, the products were double digested by appropriate restriction enzymes listed in Table 2 (bold type). The digested DNA fragments were purified using an electro-eluction method (Sambrook et al (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) and ligated into the corresponding cloning site on pGEX2T plasmid DNA (Pharmacia) by T4 DNA ligase (NEB)

at 16°C overnight. The resulting recombinant plasmid DNA was amplified in E. coli XL1-blue cells and was confirmed with the restriction enzymes listed in Table 2 (bold type) following the method described by Sambrook et al (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Habor Laboratory, Cold Spring Harbor, NY).

Preparation of glucosyltransferase recombinant proteins

E. coli cells carrying recombinant plasmid DNA as described above were grown at 37°C overnight on 2YT (16 g bacto tryptone, 10 g bacto-yeast extract, 5 g NaCl per litre) agar (1.8% w/v) plate which contained 50 μg/ml ampicillin. A single colony was picked into 2 ml of 2YT containing the same concentration of ampicillin. The bacterial culture was incubated at 37°C with moderate agitation for 6h. The bacterial culture was transferred into 1 L of 2YT and incubated at 20°C subsequently. 0.1 mM IPTG was added when the culture reached logarithmic growth phase ($A_{600\ nm} \sim 0.5$). The bacterial culture was incubated for another 24 h. The cells were collected by a centrifugation at 7,000 x g for 5 min at 4°C and resuspended in 5 ml spheroblast buffer (0.5 mM EDTA, 750 mM sucrose, 200 mM Tris, pH 8.0). Lysozyme solution was added to a final concentration of 1 mg/ml. 7-fold volume of 0.5 x spheroblast buffer was poured into the suspension immediately and the suspension was incubated for 4°C for 30 min under gentle shaking. The spheroblasts were collected by a centrifugation at 12,000 x g for 5 min at 4°C, and resuspended in 5 ml ice cold PBL buffer (140 mM NaCl, 80 mM, $NA_2\ HPO_4$, 15 mM KH_2PO4). 2 mM of PMSF was added into suspension immediately and the suspension was centrifuged at 12,000 x g for 20 min at 4°C in order to remove the cell debris. After the centrifugation, the supernatant was transferred to a 15 ml tube. 200 μl of 50% (v/v) slurry of Glutathione-coupled Sepharose 4B were added into the tube and the mixture was mixed gently for 30 min at room temperature. The mixture was then centrifuged at a very slow speed (500 x g) for 1 min. the supernatant was discarded. The beads were washed with 5 ml ice cold PBS buffer three times. After each wash, a short centrifugation was applied as described above to sediment the Sepharose beads. To recover the expressed protein from Sepharose beads, 100 μl of 20 mM reduced glutathione were used to resuspend the beads. After 10 min incubation at room temperature, the beads were collected and the supernatant containing the expressed protein was collected. The elution step was repeated once, and both supernatant fractions were combined and stored at 4°C for protein assay and further studies.

Protein concentration assay

The protein assays were carried out by adding 10 µl of protein solution into 900 µl of distilled water and 200 µl of Bio-Rad Protein Assay Dye. The absorbance at 595 nm was read. A series of BSA (bovine serium albumin) at different concentration was used as standard. Regression line was plotted based on the coordinates of the BSA concentration against the reading at 595 nm. The concentration of protein sample was therefore estimated from the regression line after the protein assay.

Assay for enzyme activity

A standard glucosylation reaction was set up by mixing 2 µg of recombinant proteins with 14 mM 2-mercaptoethanol, 5 mM UDP-glucose, 1mM of various lignin or antioxidant substrate, 100 mM Tris, pH 7.0, to a total volume of 200 µl. The reaction was carried out at 30°C for 30 min and stopped by the addition of 20 µl trichloroacetic acid (240 mg/ml). All the samples were stored at -20°C before the liquid chromatographic assay.

High-Performance Liquid Chromatographic

Reverse-phase HPLC (Waters Separator 2690 and Waters Tunable Absorbance Detector 486, Waters Limited, Herts, UK) using a Columbus 5 µ C_{18} column (250 × 4.60 mm, Phenomenex). Linear gradient of acetonitrile in H_2O (all solutions contained 0.1% trifluoroacetic acid) at 1 ml/min over 20 min, was used to separate the glucose conjugates from their aglycone. The HPLC methods were described as the following: cinnamic acid, $\lambda_{288\ nm}$, 10-55% acetonitrile; *p*-coumaric acid, $\lambda_{311\ nm}$, 10-25% acetonitrile; caffeic acid, $\lambda_{311\ nm}$, 10-16% acetonitrile; ferulic acid, $\lambda_{311\ nm}$, 10-35% acetonitrile; sinapic acid, $\lambda_{306\ nm}$, 10-40% acetonitrile; *p*-coumaryl aldehyde, $\lambda_{315\ nm}$, 10-46% acetonitrile; coniferyl aldehyde, $\lambda_{283\ nm}$, 10-47% acetonitrile; sinapyl aldehyde, $\lambda_{280\ nm}$, 10-47% acetonitrile; *p*-coumaryl alcohol, $\lambda_{283\ nm}$, 10-27% acetonitrile; coniferyl alcohol, $\lambda_{306\ nm}$, 10-25% acetonitrile; sinapyl alcohol, $\lambda_{285\ nm}$, 10-25% acetonitrile. The retention time (R_t) of the glucose conjugates analysed is listed in the following: cinnamoylglucose, R_t = 12.3 min; *p*-coumaroylglucose, R_t = 10.6 min; caffeoylglucose, R_t = 8.5 min; feruloylglucose, R_t = 10.3 min; sinapoylglucose, R_t = 9.7 min; caffeoyl-4-*O*-glucoside, R_t = 6.8 min; feruloyl-4-*O*-glucoside, R_t = 7.8 min; sinapoyl-4-*O*-glucoside, R_t = 8.2 min; coniferin, R_t = 8.2 min; syringin, R_t = 9.1 min.

The recombinant GTases were shown to have GTase activity towards the major intermediates of the lignin biosynthesis pathway (Tables 5 and 6). It is clear from these results that the GTases display different specific activity reaction profiles relative to each other on the various lignin precursor substrates utilised.

Michaelis-Menten kinetics were also studied on several of the GTases against their preferred substrates (Tables 7 and 8). It is clear from these results that the GTases display different enzyme kinetics for different substrates.

The results (in total) indicate that certain GTases show a greater potential for use in the alteration of lignin biosynthesis in *planta* than others.

Reducing the Formation of Monolignol Glucosides *In Planta*

In one approach to reduce the formation of monolignol glucosides *in planta*, A119 and A407 are down regulated using an antisense strategy (A). Expression of the A119 and A407 antisense sequences is driven by the gene's own promoter. An alternative approach (B) is to modify the UDP-glucose binding motif through an *in vitro* mutagenesis method (Lim et al., 1998) such that the mutant protein is able to bind the monolignol substrates but loses its catalytic activity. Such mutant proteins are thought to compete with the functional native protein by binding specifically to monolignols. thereby reducing the formation of monolignol glucosides.

Anti-sense approach

Amplification and cloning of the A119 and A407 promoter sequences

Approximately 2kb of the 5' flanking sequences of A119 and A407 are amplified directly from genomic DNA by PCR. The promoter fragments are then cloned into a pBluescript plasmid vector (Sambrook et al., 1989).

Construct chimaeric genes of A119 and A407 promoter and their ORF region in antisense orientation.

The A119 and A407 cDNA fragments are amplified from pGEXA119 and pGEXA407 by PCR. The fragments are then ligated correspondingly into the A119 and A407 promoter constructs described in (A)-(1) with the ORF region in the antisense orientation.

Preparation of binary construct containing the A119 and A407 antisense chimaeric gene

The DNA fragments containing the A119 and A407 antisense chimaeric genes are amplified by PCR from the chimaeric constructs described in (A)-(2). The fragments are then ligated into a binary vector (Sambrook et al., 1989). The final constructs are transformed into plants subsequently.

Mutant gene approach

***In vitro* site mutagenesis to modify the UDPglucose binding motif in A119 and A407**

In vitro site mutagenesis is carried by PCR to modify the sequences encoding the UDPglucose binding motif in A119 and A407 (Lim et al., 1998). The constructs pGEXA119 and pGEXA407 are used in the DNA templates in the PCR reaction.

Construct chimaeric mutant genes regulated by A119 and A407 promoters

The A119 and A407 mutant genes are amplified from the pGEXA119 and pGEXA407 mutant constructs described in (B)-(1) by PCR. The A119 and A407 mutant gene fragments are then ligated into the A119 and A407 promoter constructs described in (A)-(1) with the ORF region in the sense orientation.

Preparation of binary construct containing the chimaeric mutant genes A119 and A407

The DNA fragments containing the A119 and A407 mutant chimaeric genes are amplified by PCR from the chimaeric constructs described in (B)-(2). The fragments are then ligated into a binary vector (Sambrook et al., 1989). The final constructs are transformed into plants subsequently.

Enhancing the formation of monolignol glucosides *in planta*

The CaMV 35S promoter fragment is used to drive the expression of A119 and A497. DNA fragments containing A119 and A407 ORF sequences are amplified from pGEXA119 and pGEXA407 correspondingly by PCR. The DNA fragments are ligated downstream of the

CaMV 35S promoter. The constructs are used to transform plants such that the lignin content and composition is altered.

Table 1 Database information on eleven *Arabidopsis* GTase genes

Gene name	protein_id	chromosome	Database acc. no.	BAC/P1 clone	gene name in database
A062	Gi|3935156	I	ac005106	T25N20	T25N20.20
A320	Not annotated	III	ab019232	MIL23	not annotated
A41	Emb|CAB10326.1	IV	z97339	FCA4	d13780c
A42	Emb|CAB10327.1	IV	z97339	FCA4	d13785c
A43	Emb|CAB10328.1	IV	z97339	FCA4	d13790c
A911	Gi|2642451	II	ac002391	T20D16	T20D16.11
A119	Not annotated	V	ab018119	MSN2	not annotated
A233	Wrongly annotated	IV	al021961	F28A23	wrongly annotated
A407	Gi|3319344	V	af077407	F9D12	F9D12.4
A961	Gi|3582329	II	ac005496	T27A16	T27A16.15
A962	Gi|3582341	II	ac005496	T27A16	T27A16.16

Parameters used for the search of the above *Arabidopsis* sequences and the programme used are as follows:

NETBLAST with the default settings:

Infile2=nr

Matrix=Blosum 62

Translate=1

Dbtranslate=1

Table 2 DNA sequences and restriction enzyme sites in primers used in amplification of 11 *Arabidopsis* Gtase sequences from genomic DNA.

Sequence complementary to either end of the ORFs are underlined. Restriction enzyme sites that were used in making expression constructs were in BOLD type.

primer	DNA sequence (5'→3')	restriction enzyme sites
A062 5'	CGGGTGATCAGGTACCATGGCGCCACCGCATTTTC	BclI and KpnI
A062 3'	CGGAATTCGTCGACTTACTTTACTTTTACCTCCTC	EcoRI and SalI
A320 5'	CCCCCGGGTACCATGGAGCTAGAATCTTCTCTCC	SmaI and KpnI
A320 3'	CGGAATTCTCGAGTTAAAAGCTTTTGATTGATCC	EcoRI and XhoI
A41 5'	TGGGATCCATATCAGAAATGGTGTTC	BamHI
A41 3'	GGGAATTCCTAGTATCCATTATCTTTAGTC	EcoRI
A42 5'	GGGGATCCATGGACCCGTCTCGTCATACTC	BamHI
A42 3'	GGGAATTCCACTAGTGTTCTCCGTTGTCTTC	EcoRI
A43 5'	GGGGATCCAATATGGAGATGGAATCGTCGTTAC	BamHI
A43 3'	GGGAATTCCTTACACGACATTATTAATGTTTG	EcoRI
A911 5'	GGGGTACCTGATCAATAATGGGCAGTAGTGAGGG	KpnI and BclI
A911 3'	CGGAATTCGTCGACGAGTTAGGCGATTGTGATATC	EcoRI and SalI
A119 5'	CGGGATCCGGTACCATGCATATCACAAAACCACAC	BamHI and KpnI
A119 3'	CGGAATTCGCTAGCTAAGCACCACGTGACAAGTCC	EcoRI and NheI
A233 5'	CGGGATCCGGTACCATGAGTAGTGATCCTCATCGT	BamHI and KpnI
A233 3'	CGGGATCCGAATTCTACGAGGTAAACTCTTCTATG	BamHI and EcoRI

A407 5'	CGGGATCCGGTACCATGCATATCACAAAACCAC	BamHI and KpnI
A407 3'	CGGAATTCGTCGACCTAAGCACCACGTCCCAAG	EcoRI and SalI
A961 5'	GGGTGATCAGGTACCATGGGGAAGCAAGAAGATG	BclI and KpnI
A961 3'	CGGAATTCGTCGACTACTTACTTATAGAAACGCCG	EcoRI and SalI
A962 5'	GAAGATCTGGTACCATGGCGAAGCAGCAAGAAG	BglII and KpnI
A962 3'	CGGAATTCGTCGACCGATCAAAGCCCATCTATG	EcoRI and SalI

Table 3 PCR programme

Stage I (1 cycle)		Stage II (40 cycles)		Stage III (1 cycle)	
95°C	5 min	95°C	1 min	95°C	2 min
55°C	2 min	55°C	1 min	55°C	2 min
72°C	3 min	72°C	2 min	72°C	5 min

Table 4 The HPLC conditions

Lignin Precursors	Acetonitrile Gradient (%)	Detector Wavelength (nm)
cinnamic acid	10-55	288
p-coumaric acid	10-25	311
caffeic acid	10-16	311
ferulic acid	10-35	311
sinapic acid	10-40	306
p-coumaryl aldehyde	10-46	315
coniferyl aldehyde	10-47	283
sinapyl aldehyde	10-47	280
p-coumaryl alcohol	10-27	283
coniferyl alcohol	10-25	306
sinapyl alcohol	10-25	285

Table 5 Specific activity of the recombinant GTases producing glucose ester against lignin precursors

Each assay contained 0.5 m□ of potential substrates, 5 mM UDPG and 0.2 µg of recombinant GTases in a total volume of 200 µl. The reactions were incubated at 20 °C for 30 min and were stopped by addition of 20 µl TCA (240 mg/ml). Each reaction mix was then analysed using HPLC. The specific activity (nkat/mg) of the recombinant GTase is defined as the amount of substrate (nmole) converted to glucose ester per second by 1 mg of protein at 20 °C under the reaction conditions.

	A41	A320	A42	A43	A911	A062
Cinnamic acid	0.30	0.06	14.21	0.02	8.77	1.62
p-coumaric acid	13.53	0.05	4.69	0.03	4.31	2.54
Caffeic acid	2.61	0.05	0.62	0.01	0.77	0.26
Ferulic acid	6.64	0.54	15.63	0.04	2.88	0.08
Sinapic acid	5.35	15.58	11.97	0.05	0.15	0.1

Table 6 Specific activity of the recombinant GTases producing O-glucosides against lignin precursors

The reactions were set up following the conditions described in Table 1. All the reactions, except those containing the aldehydes, were stopped by the addition of TCA. The aldehyde assay mixs were injected into HPLC immediately after the reactions were completed. The specific activity (nkat/mg) of the recombinant GTase is defined as the amount of substrate (nmole) converted to 4-O-glucoside per second by 1 mg of protein at 30 °C under the reaction conditions.

	A233	A119	A407	A961	A962
Cinnamic acid	ND[a]	ND	ND	ND	ND
p-coumaric acid	0.09	0.02	0.01	0.01	0.01
caffeic acid	0.48	0.13	0.07	0.07	ND
ferulic acid	0.37	14.48	0.25	ND	ND
sinapic acid	0.39	102.56	65.39	0.01	0.01
p-coumaryl aldehyde	ND	0.03	ND	0.01	0.02
Coniferyl aldehyde	ND	1.08	ND	0.16	0.34
sinapyl aldehyde	ND	4.55	ND	0.57	0.50
p-coumaryl alcohol	ND	ND	ND	ND	ND
Coniferyl alcohol	0.46	67.53	2.78	0.57	0.49
sinapyl alcohol	0.05	126.16	114.76	0.35	0.45

[a]ND, not detected

Table 7 Kinetic studies on the recombinant GTases producing glucose esters against lignin precursors

A41		A320		A42		A911		A062	
K_m	V_{max}	K_m	V_{max}	K_m	V_{max}	K_m	V_{max}	K_m	V_{max}
MM	nkat/mg	mM	nkat/mg	mM	nkat/mg	mM	nkat/mg	mM	nkat/mg
1.51	—	1.80	—	0.72	—	1.05	—	2.36	—
—	—	—	—	0.49	19.42	0.05	9.06	4.33	2.87
0.10	16.13	—	—	0.40	6.67	0.39	11.10	5.05	4.91
0.06	20.24	—	—	0.20	1.67	0.23	1.18	—	—
0.35	11.35	—	—	0.36	18.35	0.34	6.91	—	—
0.24	6.78	0.06	8.37	0.13	12.80	—	—	—	—

Table 8 Kinetic studies on the recombinant GTases producing O-glucosides against lignin precursors

	A119		A407	
	K_m	V_{max}	K_m	V_{max}
	mM	nkat/mg	mM	nkat/mg
UDPG	0.93	—	0.89	—
ferulic acid	0.25	18.87	—	—
sinapic acid	0.51	131.58	0.14	75.19
coniferyl alcohol	0.26	92.59	—	—
sinapyl alcohol	1.10	322.58	1.07	357.10

Table 9

^{1}H and ^{13}C NMR spectra were recorded in deuterated methanol at 500 MHz and 125 MHz respectively. Chemical shifts are given on δ scale with TMS as internal standard. The position on the aromatic ring begins with the carbon joining the propanoic acid. *d*, doublet; *dd*, doublet of doublets; *m*, multiplet; *J*, coupling constant.

Position	Caffeic acid		Caffeoyl-3-*O*-glucoside	
	δ_H	δ_C	δ_H	δ_C
C1	—	128.1	—	127.6
C2	7.02 (1H, *d*, *J* = 2.0 Hz)	115.2	7.47 (1H, *d*, *J* = 2.0 Hz)	117.0
C3	—	146.7	—	146.0
C4	—	149.4	—	150.6
C5	6.77 (1H, *d*, *J* = 8.0 Hz)	116.6	6.84 (1H, *d*, *J* = 8.5 Hz)	117.8
C6	6.92 (1H, *dd*, *J* = 8.0, 2.0 Hz)	122.8	7.13 (1H, *dd*, *J* = 8.5, 2.0 Hz)	125.6
C7	7.53 (1H, *d*, *J* = 16.0 Hz)	146.9	7.45 (1H, *d*, *J* = 14.5 Hz)	146.6
C8	6.21 (1H, *d*, *J* = 15.5 Hz)	116.3	6.33 (1H, *d*, *J* = 14.5 Hz)	116.1
C9	—	171.5	—	170.4
Glc-1			~4.86 (signal interrupted)	103.9
Glc-2			3.40-3.50 (4H, *m*)	74.5
Glc-3				78.0
Glc-4				71.0
Glc-5				77.2
Glc-6			3.93 (1H, *dd*, *J* = 12.0, 2.0 Hz) 3.71 (1H, *dd*, *J* = 12.0, 5.5 Hz)	62.4

Table 10

Each assay contained 1 □g of UGT71C1, 1 mM phenolic compound, 5 mM UDP-glucose, 1.4 mM 2-mercaptoethanol and 50 mM TRIS-HCl, pH 7.0. The mix was incubated at 30 °C for 30 min. The reaction was stopped by the addition of 20 □l of trichloroacetic acid (240 mg/ml) and was analysed by reverse-phase HPLC subsequently. The results represent the mean of three replicates ± standard deviation.

Substrate	Specific activity
	nkat/mg
o-Coumaric acid	1.5 ± 0.2
m-Coumaric acid	1.2 ± 0.2
p-Coumaric acid	0
Caffeic acid	2.9 ± 0.8
Ferulic acid	0
Sinapic acid	0
Esculetin	34.8 ± 4.2
Scopoletin	29.4 ± 3.9
Salicylic acid	0
4-hydroxybenzoic acid	0
3,4-dihydroxybenzoic acid	0
Eriodictyol	0
Luteolin	0.7 ± 0.1
Quercetin	1.4 ± 0.4
Catechin	0
Cyanidin	0

CLAIMS

1. A transgenic plant comprising a nucleic acid molecule which encodes a polypeptide which:

i) has GTase activity;

v) is selected from the group comprising: sequences of Figures 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A, 10A , 11A, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32.

vi) nucleic acids which hybridise to the sequences represented in (ii) above; and

vii) nucleic acid sequences which are degenerate as a result of the genetic code to the sequences defined in (ii) and (iii) above.

2. A transgenic plant according to Claim 1 wherein the nucleic acid molecule anneals under stringent hybridisation conditions to the sequence presented in Figures 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A, 10A, 11A, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32.

3. A transgenic plant according to Claim 1 or 2 wherein the nucleic acid molecule is selected from Figure 7A or 9A.

4. A transgenic plant comprising a nucleic acid molecule wherein the nucleic acid molecule is selected from the group comprising:

i) antisense sequences selected from the group comprising: sequences of Figures 1C, 2C, 3C, 4C, 5C, 6C, 7C, 8C, 9C, 10C and 11C or parts thereof, or antisense sequences of the sense sequences presented in Figs 12 - 32;

ii) antisense sequences which will hybridise to the sense sequences according to any of Claims 1 –3 and which inhibit GTase activity.

5. A transgenic plant according to Claim 4 wherein the antisense sequence is selected from Figure 7C or 9C.

6. A transgenic plant according to any of Claims 1-5 wherein the nucleic acid is cDNA.

7. A transgenic plant according to any of Claims 1-5 wherein the nucleic acid is genomic DNA.

8. A transgenic plant according to any of Claims 1-7 wherein the plant is a woody plant selected from: poplar; eucalyptus; Douglas fir; pine; walnut; ash; birch; oak; teak; spruce.

9. A transgenic plant according to Claim 8 wherein the woody plant is a plant used typically in the paper industry, for example poplar.

10. A transgenic plant according to any of Claims 1-7 wherein the plant is a non-woody plant species.

11 A method for the manufacture of paper or board comprising:

i) pulping transgenic wood material derived from a transgenic woody plant according to any of Claims 1-10; and

ii) producing paper from said pulped transgenic wood material.

12. Paper having the characteristics of paper manufactured by the method according to Claim 11.

13. A product comprising the paper according to Claim 12.

14. A transgenic eukaryotic cell comprising a nucleic acid molecule which encodes a polypeptide which:

i) has GTase activity;

ii) is selected from the group comprising sequences of Figures 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A, 10A, 11A, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32;

iii) nucleic acids which hybridise to the sequences represented in (ii) above; and

iv) nucleic acid sequences which are degenerate as a result of the genetic code to the sequences defined in (ii) and (iii) above.

15 A transgenic eukaryotic cell according to Claim 14 wherein the nucleic acid sequences is selected from: Figures 1A, 3A, 4A, 5A, 8A, 10A, 11A, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32.

16 A transgenic eukaryotic cell according to Claim 15 wherein the nucleic acid sequence is presented in Figure 1A, 3A, 4A, 5A,7A, 8A, 9A, 10A.

17 Use of the eukaryotic cell according to any of Claims 14 -16 for the glucosylation of: caffeic acid; luteolin; quercitin; catechin; syadinin.

18. A transgenic prokaryotic cell comprising a nucleic acid molecule which encodes a polypeptide which:

i) has GTase activity;

ii) is selected from the group comprising sequences of Figures 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8A, 9A, 10A, 11A, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32;

iii) nucleic acids which hybridise to the sequences represented in (ii) above; and nucleic acid sequences which are degenerate as a result of the genetic code to the sequences defined in (ii) and (iii) above.

19 A transgenic prokaryotic cell according to Claim 18 wherein the nucleic acid sequences is selected from: Figures 1A, 3A, 4A, 5A, 8A, 10A, 11A, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 29, 30, 31, 32.

20 A transgenic prokaryotic cell according to Claim 19 wherein the nucleic acid sequence is presented in Figure 1A, 3A, 4A, 5A,7A, 8A, 9A, 10A.

21. Use of the prokaryotic cell according to any of Claims 18 – 20 for the glucosylation of: caffeic acid; luteolin; quercitin; catechin; syadinin.

FIGURE 1A **A062 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGGCGCCAC CGCATTTTCT ACTGGTAACG TTTCCGGCGC AAGGTCACGT
  51  GAACCCATCT CTCCGTTTTG CTCGTCGGCT CATCAAAAGA ACCGGCGCAC
 101  GTGTCACTTT CGTCACTTGT GTCTCCGTCT TCCACAACTC CATGATCGCA
 151  AACCACAACA AAGTCGAAAA TCTCTCTTTC CTTACTTTCT CCGACGGTTT
 201  CGACGATGGA GGCATTTCCA CCTACGAAGA CCGTCAGAAA AGGTCGGTGA
 251  ATCTCAAGGT TAACGGCGAT AAGGCACTAT CGGATTTCAT CGAAGCTACT
 301  AAGAATGGTG ACTCTCCCGT GACTTGCTTG ATCTACACGA TTCTTCTCAA
 351  TTGGGCTCCA AAAGTAGCAC GTAGATTTCA ACTTCCCTCC GCTCTTCTCT
 401  GGATCCAACC GGCTTTGGTT TTCAACATCT ATTACACTCA TTTCATGGGA
 451  AACAAGTCCG TTTTCGAGTT ACCTAATCTG TCTTCTCTGG AAATCAGAGA
 501  TCTTCCATCT TTCCTCACAC CTTCCAACAC AAACAAAGGC GCATACGATG
 551  CGTTTCAAGA AATGATGGAG TTTCTCATAA AAGAAACCAA ACCGAAAATT
 601  CTCATCAACA CTTTCGATTC GCTGGAACCA GAGGCCTTAA CGGCTTTCCC
 651  GAATATCGAT ATGGTGGCGG TTGGTCCTTT ACTTCCCACG GAGATTTTCT
 701  CAGGAAGCAC CAACAAATCA GTTAAAGATC AAAGTAGTAG TTATACACTT
 751  TGGCTAGACT CGAAAACAGA GTCCTCTGTT ATTTACGTTT CCTTTGGAAC
 801  AATGGTTGAG TTGTCCAAGA AACAGATAGA GGAACTAGCG AGAGCACTCA
 851  TAGAAGGGAA ACGACCGTTT TTGTGGGTTA TAACTGATAA ATCCAACAGA
 901  GAAACGAAAA CAGAAGGAGA AGAAGAGACA GAGATTGAGA AGATAGCTGG
 951  ATTCAGACAC GAGCTTGAAG AGGTTGGGAT GATTGTGTCG TGGTGTTCGC
1001  AGATAGAGGT TTTAAGTCAC CGAGCCGTAG GTTGTTTTGT GACTCATTGT
1051  GGGTGGAGCT CGACGCTGGA GAGTTTGGTT CTTGGCGTTC CGGTTGTGGC
1101  GTTTCCGATG TGGTCGGATC AACCGACGAA CGCGAAGCTA CTGGAAGAAA
1151  GTTGGAAGAC TGGTGTGAGG GTAAGAGAGA ACAAGGATGG TTTGGTGGAG
1201  AGAGGAGAGA TCAGGAGGTG TTTGGAAGCC GTGATGGAGG AGAAGTCGGT
1251  GGAGTTGAGG GAAAACGCAA AGAAATGGAA GCGTTTAGCG ATGGAAGCGG
1301  GTAGAGAAGG AGGATCTTCG GATAAGAACA TGGAGGCTTT TGTGGAGGAT
1351  ATTTGTGGAG AATCTCTTAT TCAAAACTTG TGTGAAGCAG AGGAGGTAAA
1401  AGTAAAGTAA
```

FIGURE 1B **A062 AMINO ACID SEQUENCE**

```
  1  MAPPHFLLVT FPAQGHVNPS LRFARRLIKR TGARVTFVTC VSVFHNSMIA
 51  NHNKVENLSF LTFSDGFDDG GISTYEDRQK RSVNLKVNGD KALSDFIEAT
101  KNGDSPVTCL IYTILLNWAP KVARRFQLPS ALLWIQPALV FNIYYTHFMG
151  NKSVFELPNL SSLEIRDLPS FLTPSNTNKG AYDAFQEMME FLIKETKPKI
201  LINTFDSLEP EALTAFPNID MVAVGPLLPT EIFSGSTNKS VKDQSSSYTL
251  WLDSKTESSV IYVSFGTMVE LSKKQIEELA RALIEGKRPF LWVITDKSNR
301  ETKTEGEEET EIEKIAGFRH ELEEVGMIVS WCSQIEVLSH RAVGCFVTHC
351  GWSSTLESLV LGVPVVAFPM WSDQPTNAKL LEESWKTGVR VRENKDGLVE
401  RGEIRRCLEA VMEEKSVELR ENAKKWKRLA MEAGREGGSS DKNMEAFVED
451  ICGESLIQNL CEAEEVKVK
```

FIGURE 1C **A062 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  TTACTTTACT TTTACCTCCT CTGCTTCACA CAAGTTTTGA ATAAGAGATT
  51  CTCCACAAAT ATCCTCCACA AAAGCCTCCA TGTTCTTATC CGAAGATCCT
 101  CCTTCTCTAC CCGCTTCCAT CGCTAAACGC TTCCATTTCT TTGCGTTTTC
 151  CCTCAACTCC ACCGACTTCT CCTCCATCAC GGCTTCCAAA CACCTCCTGA
 201  TCTCTCCTCT CTCCACCAAA CCATCCTTGT TCTCTCTTAC CCTCACACCA
 251  GTCTTCCAAC TTTCTTCCAG TAGCTTCGCG TTCGTCGGTT GATCCGACCA
 301  CATCGGAAAC GCCACAACCG GAACGCCAAG AACCAAACTC TCCAGCGTCG
 351  AGCTCCACCC ACAATGAGTC ACAAAACAAC CTACGGCTCG GTGACTTAAA
 401  ACCTCTATCT GCGAACACCA CGACACAATC ATCCCAACCT CTTCAAGCTC
 451  GTGTCTGAAT CCAGCTATCT TCTCAATCTC TGTCTCTTCT TCTCCTTCTG
 501  TTTTCGTTTC TCTGTTGGAT TTATCAGTTA TAACCCACAA AAACGGTCGT
 551  TTCCCTTCTA TGAGTGCTCT CGCTAGTTCC TCTATCTGTT TCTTGGACAA
 601  CTCAACCATT GTTCCAAAGG AAACGTAAAT AACAGAGGAC TCTGTTTTCG
 651  AGTCTAGCCA AAGTGTATAA CTACTACTTT GATCTTTAAC TGATTTGTTG
 701  GTGCTTCCTG AGAAAATCTC CGTGGGAAGT AAAGGACCAA CCGCCACCAT
 751  ATCGATATTC GGGAAAGCCG TTAAGGCCTC TGGTTCCAGC GAATCGAAAG
 801  TGTTGATGAG AATTTTCGGT TTGGTTTCTT TTATGAGAAA CTCCATCATT
 851  TCTTGAAACG CATCGTATGC GCCTTTGTTT GTGTTGGAAG GTGTGAGGAA
 901  AGATGGAAGA TCTCTGATTT CCAGAGAAGA CAGATTAGGT AACTCGAAAA
 951  CGGACTTGTT TCCCATGAAA TGAGTGTAAT AGATGTTGAA AACCAAAGCC
1001  GGTTGGATCC AGAGAAGAGC GGAGGGAAGT TGAAATCTAC GTGCTACTTT
1051  TGGAGCCCAA TTGAGAAGAA TCGTGTAGAT CAAGCAAGTC ACGGGAGAGT
1101  CACCATTCTT AGTAGCTTCG ATGAAATCCG ATAGTGCCTT ATCGCCGTTA
1151  ACCTTGAGAT TCACCGACCT TTTCTGACGG TCTTCGTAGG TGGAAATGCC
1201  TCCATCGTCG AAACCGTCGG AGAAAGTAAG GAAAGAGAGA TTTTCGACTT
1251  TGTTGTGGTT TGCGATCATG GAGTTGTGGA AGACGGAGAC ACAAGTGACG
1301  AAAGTGACAC GTGCGCCGGT TCTTTTGATG AGCCGACGAG CAAAACGGAG
1351  AGATGGGTTC ACGTGACCTT GCGCCGGAAA CGTTACCAGT AGAAAATGCG
1401  GTGGCGCCAT
```

FIGURE 2A **A320 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGGAGCTAG AATCTTCTCC TCCTCTACCT CCTCATGTGA TGCTCGTATC
  51  TTTTCCAGGG CAAGGCCACG TTAATCCACT TCTTCGTCTT GGTAAGCTCT
 101  TAGCTTCAAA GGGTTTGCTC ATAACCTTCG TCACCACTGA GTCATGGGGC
 151  AAAAAGATGC GAATCTCCAA CAAAATCCAA GACCGTGTCC TCAAACCGGT
 201  TGGTAAAGGC TATCTCCGGT ATGATTTCTT CGACGACGGG CTTCCTGAAG
 251  ACGACGAAGC TAGCAGAACC AACTTAACCA TCCTCCGACC ACATCTAGAG
 301  CTGGTCGGCA AAAGAGAGAT CAAGAACCTT GTGAAACGTT ACAAGGAAGT
 351  AACGAAACAG CCCGTGACAT GTCTTATCAA CAACCCTTTC GTCTCTTGGG
 401  TCTGTGACGT GGCAGAAGAT CTTCAAATCC CTTGTGCTGT TCTTTGGGTT
 451  CAATCTTGTG CCTGCTTAGC TGCTTATTAC TATTACCACC ACAACCTAGT
 501  TGACTTCCCG ACCAAAACAG AACCCGAGAT CGATGTCCAA ATCTCTGGCA
 551  TGCCTCTCTT GAAACATGAC GAGATCCCTT CTTTCATTCA CCCTTCAAGT
 601  CCTCACTCCG CTTTGCGAGA AGTGATCATA GATCAGATTA AACGGCTTCA
 651  CAAGACTTTC TCCATTTTCA TCGACACTTT CAACTCATTG GAGAAAGACA
 701  TCATTGACCA CATGTCGACG CTCTCTCTCC CCGGTGTTAT CAGACCGCTA
 751  GGACCACTCT ACAAAATGGC TAAAACCGTA GCTTATGATG TCGTTAAAGT
 801  AAACATCTCT GAGCCAACGG ATCCTTGCAT GGAGTGGTTA GACTCGCAGC
 851  CAGTTTCCTC CGTTGTTTAC ATCTCATTCG GGACCGTTGC TTACTTGAAA
 901  CAAGAACAAA TAGACGAGAT CGCTTACGGT GTGTTAAACG CCGACGTTAC
 951  GTTCTTGTGG GTGATTAGAC AACAAGAGTT AGGTTTCAAC AAAGAGAAAC
1001  ATGTTTTGCC GGAAGAAGTT AAAGGGAAAG GGAAGATCGT TGAATGGTGT
1051  TCACAAGAGA AAGTATTATC TCATCCTTCA GTGGCATGTT TCGTGACTCA
1101  CTGTGGATGG AACTCAACGA TGGAAGCTGT GTCTTCCGGA GTCCCGACGG
1151  TTTGTTTTCC TCAATGGGGA GATCAAGTCA CGGACGCCGT TTACATGATC
1201  GATGTTTGGA AGACGGGAGT GAGGCTAAGC CGTGGAGAGG CGGAGGAGAG
1251  GTTAGTGCCG AGGGAGGAAG TTGCGGAGAG GTTGAGAGAG GTTACTAAAG
1301  GAGAGAAAGC GATCGAGTTG AAAAAGAATG CTTTGAAGTG GAAGGAAGAG
1351  GCGGAGGCGG CGGTTGCTCG CGGTGGTTCG TCGGATAGGA ATCTTGAAAA
1401  GTTTGTGGAG AAGTTGGGTG CCAAACCTGT GGGGAAAGTA CAAAACGGGA
1451  GTCATAATCA TGTCTTGGCT GGATCAATCA AAAGCTTTTA A
```

FIGURE 2B **A320 AMINO ACID SEQUENCE**

```
   1  MELESSPPLP PHVMLVSFPG QGHVNPLLRL GKLLASKGLL ITFVTTESWG
  51  KKMRISNKIQ DRVLKPVGKG YLRYDFFDDG LPEDDEASRT NLTILRPHLE
 101  LVGKREIKNL VKRYKEVTKQ PVTCLINNPF VSWVCDVAED LQIPCAVLWV
 151  QSCACLAAYY YYHHNLVDFP TKTEPEIDVQ ISGMPLLKHD EIPSFIHPSS
 201  PHSALREVII DQIKRLHKTF SIFIDTFNSL EKDIIDHMST LSLPGVIRPL
 251  GPLYKMAKTV AYDVVKVNIS EPTDPCMEWL DSQPVSSVVY ISFGTVAYLK
 301  QEQIDEIAYG VLNADVTFLW VIRQQELGFN KEKHVLPEEV KGKGKIVEWC
 351  SQEKVLSHPS VACFVTHCGW NSTMEAVSSG VPTVCFPQWG DQVTDAVYMI
 401  DVWKTGVRLS RGEAEERLVP REEVAERLRE VTKGEKAIEL KKNALKWKEE
 451  AEAAVARGGS SDRNLEKFVE KLGAKPVGKV QNGSHNHVLA GSIKSF
```

FIGURE 2C **A320 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  TTAAAAGCTT TTGATTGATC CAGCCAAGAC ATGATTATGA CTCCCGTTTT
  51  GTACTTTCCC CACAGGTTTG GCACCCAACT TCTCCACAAA CTTTTCAAGA
 101  TTCCTATCCG ACGAACCACC GCGAGCAACC GCCGCCTCCG CCTCTTCCTT
 151  CCACTTCAAA GCATTCTTTT TCAACTCGAT CGCTTTCTCT CCTTTAGTAA
 201  CCTCTCTCAA CCTCTCCGCA ACTTCCTCCC TCGGCACTAA CCTCTCCTCC
 251  GCCTCTCCAC GGCTTAGCCT CACTCCCGTC TTCCAAACAT CGATCATGTA
 301  AACGGCGTCC GTGACTTGAT CTCCCCATTG AGGAAAACAA ACCGTCGGGA
 351  CTCCGGAAGA CACAGCTTCC ATCGTTGAGT TCCATCCACA GTGAGTCACG
 401  AAACATGCCA CTGAAGGATG AGATAATACT TTCTCTTGTG AACACCATTC
 451  AACGATCTTC CCTTTCCCTT TAACTTCTTC CGGCAAAACA TGTTTCTCTT
 501  TGTTGAAACC TAACTCTTGT TGTCTAATCA CCCACAAGAA CGTAACGTCG
 551  GCGTTTAACA CACCGTAAGC GATCTCGTCT ATTTGTTCTT GTTTCAAGTA
 601  AGCAACGGTC CCGAATGAGA TGTAAACAAC GGAGGAAACT GGCTGCGAGT
 651  CTAACCACTC CATGCAAGGA TCCGTTGGCT CAGAGATGTT TACTTTAACG
 701  ACATCATAAG CTACGGTTTT AGCCATTTTG TAGAGTGGTC CTAGCGGTCT
 751  GATAACACCG GGGAGAGAGA GCGTCGACAT GTGGTCAATG ATGTCTTTCT
 801  CCAATGAGTT GAAAGTGTCG ATGAAAATGG AGAAAGTCTT GTGAAGCCGT
 851  TTAATCTGAT CTATGATCAC TTCTCGCAAA GCGGAGTGAG GACTTGAAGG
 901  GTGAATGAAA GAAGGGATCT CGTCATGTTT CAAGAGAGGC ATGCCAGAGA
 951  TTTGGACATC GATCTCGGGT TCTGTTTTGG TCGGGAAGTC AACTAGGTTG
1001  TGGTGGTAAT AGTAATAAGC AGCTAAGCAG GCACAAGATT GAACCCAAAG
1051  AACAGCACAA GGGATTTGAA GATCTTCTGC CACGTCACAG ACCCAAGAGA
1101  CGAAAGGGTT GTTGATAAGA CATGTCACGG GCTGTTTCGT TACTTCCTTG
1151  TAACGTTTCA CAAGGTTCTT GATCTCTCTT TTGCCGACCA GCTCTAGATG
1201  TGGTCGGAGG ATGGTTAAGT TGGTTCTGCT AGCTTCGTCG TCTTCAGGAA
1251  GCCCGTCGTC GAAGAAATCA TACCGGAGAT AGCCTTTACC AACCGGTTTG
1301  AGGACACGGT CTTGGATTTT GTTGGAGATT CGCATCTTTT TGCCCCATGA
1351  CTCAGTGGTG ACGAAGGTTA TGAGCAAACC CTTTGAAGCT AAGAGCTTAC
1401  CAAGACGAAG AAGTGGATTA ACGTGGCCTT GCCCTGGAAA AGATACGAGC
1451  ATCACATGAG GAGGTAGAGG AGGAGAAGAT TCTAGCTCCA T
```

FIGURE 3A **A41 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGGGATCCA TATCAGAAAT GGTGTTCGAA ACTTGTCCAT CTCCAAACCC
  51  AATTCATGTA ATGCTCGTCT CGTTTCAAGG ACAAGGCCAC GTCAACCCTC
 101  TTCTTCGTCT CGGCAAGTTA ATTGCTTCAA AGGGTTTACT CGTTACCTTC
 151  GTTACAACGG AGCTTTGGGG CAAGAAAATG AGACAAGCCA ACAAAATCGT
 201  TGACGGTGAA CTTAAACCGG TTGGTTCCGG TTCAATCCGG TTTGAGTTCT
 251  TTGATGAAGA ATGGGCAGAG GATGATGACC GGAGAGCTGA TTTCTCTTTG
 301  TACATTGCTC ACCTAGAGAG CGTTGGGATA CGAGAAGTGT CTAAGCTTGT
 351  GAGAAGATAC GAGGAAGCGA ACGAGCCTGT CTCGTGTCTT ATCAATAACC
 401  CGTTTATCCC ATGGGTCTGC CACGTGGCGG AAGAGTTCAA CATTCCTTGT
 451  GCGGTTCTCT GGGTTCAGTC TTGTGCTTGT TTCTCTGCTT ATTACCATTA
 501  CCAAGATGGC TCTGTTTCAT TCCCTACGGA AACAGAGCCT GAGCTCGATG
 551  TGAAGCTTCC TTGTGTTCCT GTCTTGAAGA ACGACGAGAT TCCTAGCTTT
 601  CTCCATCCTT CTTCTAGGTT CACGGGTTTT CGACAAGCGA TTCTTGGGCA
 651  ATTCAAGAAT CTGAGCAAGT CCTTCTGTGT TCTAATCGAT TCTTTTGACT
 701  CATTGGAACA AGAAGTTATC GATTACATGT CAAGTCTTTG TCCGGTTAAA
 751  ACCGTTGGAC CGCTTTTCAA AGTTGCTAGG ACAGTTACTT CTGACGTAAG
 801  CGGTGACATT TGCAAATCAA CAGATAAATG CCTCGAGTGG TTAGACTCGA
 851  GGCCTAAATC GTCAGTTGTC TACATTTCGT TCGGGACAGT TGCATATTTG
 901  AAGCAAGAAC AGATCGAAGA GATCGCTCAC GGAGTTTTGA AGTCGGGTTT
 951  ATCGTTCTTG TGGGTGATTA GACCTCCACC ACACGATCTG AAGGTCGAGA
1001  CACATGTCTT GCCTCAAGAA CTTAAAGAGA GTAGTGCTAA AGGTAAAGGG
1051  ATGATTGTGG ATTGGTGCCC ACAAGAGCAA GTCTTGTCTC ATCCTTCAGT
1101  GGCATGCTTC GTGACTCATT GTGGATGGAA CTCGACAATG GAATCTTTGT
1151  CTTCAGGTGT TCCGGTGGTT TGTTGTCCGC AATGGGGAGA TCAAGTGACT
1201  GATGCAGTGT ATTTGATCGA TGTTTTCAAG ACCGGGGTTA GACTAGGCCG
1251  TGGAGCGACC GAGGAGAGGG TAGTGCCAAG GGAGGAAGTG GCGGAGAAGC
1301  TTTTGGAAGC GACAGTTGGG GAGAAGGCAG AGGAGTTGAG AAAGAACGCT
1351  TTGAAATGGA AGGCGGAGGC GGAAGCAGCG GTGGCTCCAG GAGGTTCGTC
1401  GGATAAGAAT TTTAGGGAGT TTGTGGAGAA GTTAGGTGCG GGAGTAACGA
1451  AGACTAAAGA TAATGGATAC TAG
```

FIGURE 3B **A41 AMINO ACID SEQUENCE**

```
  1  MVFETCPSPN PIHVMLVSFQ GQGHVNPLLR LGKLIASKGL LVTFVTTELW
 51  GKKMRQANKI VDGELKPVGS GSIRFEFFDE EWAEDDDRRA DFSLYIAHLE
101  SVGIREVSKL VRRYEEANEP VSCLINNPFI PWVCHVAEEF NIPCAVLWVQ
151  SCACFSAYYH YQDGSVSFPT ETEPELDVKL PCVPVLKNDE IPSFLHPSSR
201  FTGFRQAILG QFKNLSKSFC VLIDSFDSLE QEVIDYMSSL CPVKTVGPLF
251  KVARTVTSDV SGDICKSTDK CLEWLDSRPK SSVVYISFGT VAYLKQEQIE
301  EIAHGVLKSG LSFLWVIRPP PHDLKVETHV LPQELKESSA KGKGMIVDWC
351  PQEQVLSHPS VACFVTHCGW NSTMESLSSG VPVVCCPQWG DQVTDAVYLI
401  DVFKTGVRLG RGATEERVVP REEVAEKLLE ATVGEKAEEL RKNALKWKAE
451  AEAAVAPGGS SDKNFREFVE KLGAGVTKTK DNGY
```

FIGURE 3C **A41 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  CTAGTATCCA TTATCTTTAG TCTTCGTTAC TCCCGCACCT AACTTCTCCA
  51  CAAACTCCCT AAAATTCTTA TCCGACGAAC CTCCTGGAGC CACCGCTGCT
 101  TCCGCCTCCG CCTTCCATTT CAAAGCGTTC TTTCTCAACT CCTCTGCCTT
 151  CTCCCCAACT GTCGCTTCCA AAAGCTTCTC CGCCACTTCC TCCCTTGGCA
 201  CTACCCTCTC CTCGGTCGCT CCACGGCCTA GTCTAACCCC GGTCTTGAAA
 251  ACATCGATCA AATACACTGC ATCAGTCACT TGATCTCCCC ATTGCGGACA
 301  ACAAACCACC GGAACACCTG AAGACAAAGA TTCCATTGTC GAGTTCCATC
 351  CACAATGAGT CACGAAGCAT GCCACTGAAG GATGAGACAA GACTTGCTCT
 401  TGTGGGCACC AATCCACAAT CATCCCTTTA CCTTTAGCAC TACTCTCTTT
 451  AAGTTCTTGA GGCAAGACAT GTGTCTCGAC CTTCAGATCG TGTGGTGGAG
 501  GTCTAATCAC CCACAAGAAC GATAAACCCG ACTTCAAAAC TCCGTGAGCG
 551  ATCTCTTCGA TCTGTTCTTG CTTCAAATAT GCAACTGTCC CGAACGAAAT
 601  GTAGACAACT GACGATTTAG GCCTCGAGTC TAACCACTCG AGGCATTTAT
 651  CTGTTGATTT GCAAATGTCA CCGCTTACGT CAGAAGTAAC TGTCCTAGCA
 701  ACTTTGAAAA GCGGTCCAAC GGTTTTAACC GGACAAAGAC TTGACATGTA
 751  ATCGATAACT TCTTGTTCCA ATGAGTCAAA AGAATCGATT AGAACACAGA
 801  AGGACTTGCT CAGATTCTTG AATTGCCCAA GAATCGCTTG TCGAAAACCC
 851  GTGAACCTAG AAGAAGGATG GAGAAAGCTA GGAATCTCGT CGTTCTTCAA
 901  GACAGGAACA CAAGGAAGCT TCACATCGAG CTCAGGCTCT GTTTCCGTAG
 951  GGAATGAAAC AGAGCCATCT TGGTAATGGT AATAAGCAGA GAAACAAGCA
1001  CAAGACTGAA CCCAGAGAAC CGCACAAGGA ATGTTGAACT CTTCCGCCAC
1051  GTGGCAGACC CATGGGATAA ACGGGTTATT GATAAGACAC GAGACAGGCT
1101  CGTTCGCTTC CTCGTATCTT CTCACAAGCT TAGACACTTC TCGTATCCCA
1151  ACGCTCTCTA GGTGAGCAAT GTACAAAGAG AAATCAGCTC TCCGGTCATC
1201  ATCCTCTGCC CATTCTTCAT CAAAGAACTC AAACCGGATT GAACCGGAAC
1251  CAACCGGTTT AAGTTCACCG TCAACGATTT TGTTGGCTTG TCTCATTTTC
1301  TTGCCCCAAA GCTCCGTTGT AACGAAGGTA ACGAGTAAAC CCTTTGAAGC
1351  AATTAACTTG CCGAGACGAA GAAGAGGGTT GACGTGGCCT TGTCCTTGAA
1401  ACGAGACGAG CATTACATGA ATTGGGTTTG GAGATGGACA AGTTTCGAAC
1451  ACCATTTCTG ATATGGATCC CAT
```

FIGURE 4A **A42 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGGACCCGT CTCGTCATAC TCATGTGATG CTCGTATCTT TCCCCGGCCA
  51  AGGTCACGTA AACCCTCTAC TTCGTCTCGG AAAGCTCATA GCCTCTAAAG
 101  GCTTACTCGT CACCTTTGTC ACCACAGAGA AGCCATGGGG CAAGAAGATG
 151  CGTCAAGCCA ACAAGATTCA AGACGGTGTG CTCAAACCGG TCGGTCTAGG
 201  TTTCATCCGG TTTGAGTTCT TCTCTGACGG CTTCGCCGAC GACGATGAAA
 251  AAAGATTCGA CTTCGATGCC TTCCGACCAC ACCTTGAAGC TGTCGGAAAA
 301  CAAGAGATCA AGAATCTCGT TAAGAGATAT AACAAGGAGC CGGTGACGTG
 351  TCTCATAAAC AACGCTTTTG TCCCATGGGT ATGTGATGTC GCCGAGGAGC
 401  TTCACATCCC TTCGGCTGTT CTATGGGTCC AGTCTTGTGC TTGTCTCACG
 451  GCTTATTACT ATTACCACCA CCGGTTAGTT AAGTTCCCGA CCAAAACCGA
 501  GCCGGACATC AGCGTTGAAA TCCCTTGCTT GCCATTGTTA AAGCATGACG
 551  AGATCCCAAG CTTTCTTCAC CCTTCGTCTC CGTATACAGC TTTTGGAGAT
 601  ATCATTTTAG ACCAGTTAAA GAGATTCGAA AACCACAAGT CTTTCTATCT
 651  TTTCATCGAC ACTTTTCGCG AACTAGAAAA AGACATCATG GACCACATGT
 701  CACAACTTTG TCCTCAAGCC ATCATCAGTC CTGTCGGTCC GCTCTTCAAG
 751  ATGGCTCAAA CCTTGAGTTC TGACGTTAAG GGAGATATAT CCGAGCCAGC
 801  GAGTGACTGC ATGGAATGGC TTGACTCAAG AGAACCATCC TCAGTCGTTT
 851  ACATCTCCTT TGGGACTATA GCCAACTTGA AGCAAGAGCA GATGGAGGAG
 901  ATCGCTCATG GCGTTTTGAG CTCTGGCTTG TCGGTCTTAT GGGTGGTTCG
 951  GCCTCCCATG GAAGGGACAT TTGTAGAACC ACATGTTTTG CCTCGAGAGC
1001  TCGAAGAAAA GGGTAAAATC GTGGAATGGT GTCCCCAAGA GAGAGTCTTG
1051  GCTCATCCTG CGATTGCTTG TTTCTTAAGT CACTGCGGAT GGAACTCGAC
1101  AATGGAGGCT TTAACTGCCG GAGTCCCCGT TGTTTGTTTT CCGCAATGGG
1151  GAGATCAAGT GACTGATGCG GTGTACTTGG CTGATGTTTT CAAGACAGGA
1201  GTGAGACTAG GCCGCGGAGC CGCTGAGGAG ATGATTGTTT CGAGGGAGGT
1251  TGTAGCAGAG AAGCTGCTTG AGGCCACAGT TGGGGAAAAG GCGGTGGAGC
1301  TGAGAGAAAA CGCTCGGAGG TGGAAGGCGG AGGCCGAGGC CGCCGTGGCG
1351  GACGGTGGAT CATCTGATAT GAACTTTAAA GAGTTTGTGG ACAAGTTGGT
1401  TACGAAACAT GTGACGAGAG AAGACAACGG AGAACACTAG
```

FIGURE 4B **A42 AMINO ACID SEQUENCE**

```
  1  MDPSRHTHVM LVSFPGQGHV NPLLRLGKLI ASKGLLVTFV TTEKPWGKKM
 51  RQANKIQDGV LKPVGLGFIR FEFFSDGFAD DDEKRFDFDA FRPHLEAVGK
101  QEIKNLVKRY NKEPVTCLIN NAFVPWVCDV AEELHIPSAV LWVQSCACLT
151  AYYYYHHRLV KFPTKTEPDI SVEIPCLPLL KHDEIPSFLH PSSPYTAFGD
201  IILDQLKRFE NHKSFYLFID TFRELEKDIM DHMSQLCPQA IISPVGPLFK
251  MAQTLSSDVK GDISEPASDC MEWLDSREPS SVVYISFGTI ANLKQEQMEE
301  IAHGVLSSGL SVLWVVRPPM EGTFVEPHVL PRELEEKGKI VEWCPQERVL
351  AHPAIACFLS HCGWNSTMEA LTAGVPVVCF PQWGDQVTDA VYLADVFKTG
401  VRLGRGAAEE MIVSREVVAE KLLEATVGEK AVELRENARR WKAEAEAAVA
451  DGGSSDMNFK EFVDKLVTKH VTREDNGEH
```

FIGURE 4C **A42 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  CTAGTGTTCT CCGTTGTCTT CTCTCGTCAC ATGTTTCGTA ACCAACTTGT
  51  CCACAAACTC TTTAAAGTTC ATATCAGATG ATCCACCGTC CGCCACGGCG
 101  GCCTCGGCCT CCGCCTTCCA CCTCCGAGCG TTTTCTCTCA GCTCCACCGC
 151  CTTTTCCCCA ACTGTGGCCT CAAGCAGCTT CTCTGCTACA ACCTCCCTCG
 201  AAACAATCAT CTCCTCAGCG GCTCCGCGGC CTAGTCTCAC TCCTGTCTTG
 251  AAAACATCAG CCAAGTACAC CGCATCAGTC ACTTGATCTC CCCATTGCGG
 301  AAAACAAACA ACGGGGACTC CGGCAGTTAA AGCCTCCATT GTCGAGTTCC
 351  ATCCGCAGTG ACTTAAGAAA CAAGCAATCG CAGGATGAGC CAAGACTCTC
 401  TCTTGGGGAC ACCATTCCAC GATTTTACCC TTTTCTTCGA GCTCTCGAGG
 451  CAAAACATGT GGTTCTACAA ATGTCCCTTC CATGGGAGGC CGAACCACCC
 501  ATAAGACCGA CAAGCCAGAG CTCAAAACGC CATGAGCGAT CTCCTCCATC
 551  TGCTCTTGCT TCAAGTTGGC TATAGTCCCA AAGGAGATGT AAACGACTGA
 601  GGATGGTTCT CTTGAGTCAA GCCATTCCAT GCAGTCACTC GCTGGCTCGG
 651  ATATATCTCC CTTAACGTCA GAACTCAAGG TTTGAGCCAT CTTGAAGAGC
 701  GGACCGACAG GACTGATGAT GGCTTGAGGA CAAAGTTGTG ACATGTGGTC
 751  CATGATGTCT TTTTCTAGTT CGCGAAAAGT GTCGATGAAA AGATAGAAAG
 801  ACTTGTGGTT TTCGAATCTC TTTAACTGGT CTAAAATGAT ATCTCCAAAA
 851  GCTGTATACG GAGACGAAGG GTGAAGAAAG CTTGGGATCT CGTCATGCTT
 901  TAACAATGGC AAGCAAGGGA TTTCAACGCT GATGTCCGGC TCGGTTTTGG
 951  TCGGGAACTT AACTAACCGG TGGTGGTAAT AGTAATAAGC CGTGAGACAA
1001  GCACAAGACT GGACCCATAG AACAGCCGAA GGGATGTGAA GCTCCTCGGC
1051  GACATCACAT ACCCATGGGA CAAAAGCGTT GTTTATGAGA CACGTCACCG
1101  GCTCCTTGTT ATATCTCTTA ACGAGATTCT TGATCTCTTG TTTTCCGACA
1151  GCTTCAAGGT GTGGTCGGAA GGCATCGAAG TCGAATCTTT TTTCATCGTC
1201  GTCGGCGAAG CCGTCAGAGA AGAACTCAAA CCGGATGAAA CCTAGACCGA
1251  CCGGTTTGAG CACACCGTCT TGAATCTTGT TGGCTTGACG CATCTTCTTG
1301  CCCCATGGCT TCTCTGTGGT GACAAAGGTG ACGAGTAAGC CTTTAGAGGC
1351  TATGAGCTTT CCGAGACGAA GTAGAGGGTT TACGTGACCT TGGCCGGGGA
1401  AAGATACGAG CATCACATGA GTATGACGAG ACGGGTCCAT
```

FIGURE 5A **A43 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGGAGATGG AATCGTCGTT ACCTCATGTG ATGCTCGTAT CATTCCCAGG
  51  GCAAGGTCAC ATAAGCCCTC TTCTTCGTCT CGGAAAGATC ATTGCCTCTA
 101  AAGGCTTAAT CGTCACCTTT GTAACCACAG AGGAACCATT GGGCAAGAAG
 151  ATGCGTCAAG CCAACAATAT TCAAGACGGT GTGCTCAAAC CGGTCGGGCT
 201  AGGTTTTCTC CGGTTCGAGT TCTTCGAGGA TGGATTTGTC TACAAAGAAG
 251  ACTTTGATTT GTTACAAAAA TCACTTGAAG TTTCCGGAAA ACGAGAGATC
 301  AAGAATCTTG TCAAGAAATA TGAGAAGCAA CCAGTGAGAT GTCTCATAAA
 351  TAATGCCTTT GTTCCATGGG TTTGTGACAT AGCCGAGGAG CTTCAAATCC
 401  CATCAGCTGT TCTTTGGGTC CAGTCTTGTG CTTGCCTCGC CGCTTATTAC
 451  TATTACCACC ACCAGTTAGT TAAGTTTCCG ACCGAAACCG AGCCGGAAAT
 501  AACCGTTGAC GTCCCTTTCA AGCCATTAAC ATTGAAGCAT GACGAGATCC
 551  CTAGCTTTCT TCACCCTTCC TCTCCGCTGT CCTCTATAGG AGGTACCATT
 601  TTAGAGCAGA TCAAGCGACT TCACAAGCCT TTCTCTGTTC TCATCGAAAC
 651  TTTTCAAGAA CTTGAAAAAG ATACCATTGA CCACATGTCC CAGCTCTGCC
 701  CTCAAGTCAA CTTCAACCCC ATCGGTCCGC TTTTTACTAT GGCTAAAACC
 751  ATAAGGTCTG ACATCAAGGG AGACATCTCC AAGCCAGATA GTGACTGCAT
 801  AGAGTGGCTT GACTCGAGAG AACCATCCTC CGTTGTTTAC ATCTCTTTTG
 851  GGACTTTGGC TTTCTTGAAG CAAAACCAGA TCGACGAGAT TGCTCACGGC
 901  ATTCTCAACT CCGGGTTGTC CTGCTTATGG GTTTTGCGGC CTCCCTTAGA
 951  AGGCTTAGCC ATAGAACCGC ATGTCTTGCC TCTAGAGCTT GAAGAGAAAG
1001  GGAAGATTGT GGAATGGTGT CAACAAGAGA AAGTTTTGGC TCATCCTGCG
1051  GTTGCTTGCT TCTTAAGTCA CTGTGGATGG AACTCAACCA TGGAGGCTTT
1101  AACTTCAGGA GTTCCCGTTA TTTGTTTCCC GCAGTGGGGA GATCAGGTGA
1151  CAAATGCGGT GTACATGATT GATGTTTTCA AGACAGGATT GAGACTCAGC
1201  CGTGGAGCTT CCGATGAGAG GATTGTTCCA AGGGAGGAGG TTGCTGAGCG
1251  ACTGCTTGAG GCCACCGTTG GAGAGAAGGC GGTGGAGCTG AGAGAAAACG
1301  CTCGGAGGTG GAAGGAGGAG GCGGAGTCTG CCGTGGCTTA CGGTGGAACA
1351  TCGGAAAGGA ATTTTCAAGA GTTTGTTGAC AAGTTGGTTG ATGTCAAGAC
1401  AATGACAAAC ATTAATAATG TCGTGTAAGT
```

FIGURE 5B **A43 AMINO ACID SEQUENCE**

```
   1  MEMESSLPHV MLVSFPGQGH ISPLLRLGKI IASKGLIVTF VTTEEPLGKK
  51  MRQANNIQDG VLKPVGLGFL RFEFFEDGFV YKEDFDLLQK SLEVSGKREI
 101  KNLVKKYEKQ PVRCLINNAF VPWVCDIAEE LQIPSAVLWV QSCACLAAYY
 151  YYHHQLVKFP TETEPEITVD VPFKPLTLKH DEIPSFLHPS SPLSSIGGTI
 201  LEQIKRLHKP FSVLIETFQE LEKDTIDHMS QLCPQVNFNP IGPLFTMAKT
 251  IRSDIKGDIS KPDSDCIEWL DSREPSSVVY ISFGTLAFLK QNQIDEIAHG
 301  ILNSGLSCLW VLRPPLEGLA IEPHVLPLEL EEKGKIVEWC QQEKVLAHPA
 351  VACFLSHCGW NSTMEALTSG VPVICFPQWG DQVTNAVYMI DVFKTGLRLS
 401  RGASDERIVP REEVAERLLE ATVGEKAVEL RENARRWKEE AESAVAYGGT
 451  SERNFQEFVD KLVDVKTMTN INNVV
```

FIGURE 5C **A43 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  ACTTACACGA CATTATTAAT GTTTGTCATT GTCTTGACAT CAACCAACTT
  51  GTCAACAAAC TCTTGAAAAT TCCTTTCCGA TGTTCCACCG TAAGCCACGG
 101  CAGACTCCGC CTCCTCCTTC CACCTCCGAG CGTTTTCTCT CAGCTCCACC
 151  GCCTTCTCTC CAACGGTGGC CTCAAGCAGT CGCTCAGCAA CCTCCTCCCT
 201  TGGAACAATC CTCTCATCGG AAGCTCCACG GCTGAGTCTC AATCCTGTCT
 251  TGAAAACATC AATCATGTAC ACCGCATTTG TCACCTGATC TCCCCACTGC
 301  GGGAAACAAA TAACGGGAAC TCCTGAAGTT AAAGCCTCCA TGGTTGAGTT
 351  CCATCCACAG TGACTTAAGA AGCAAGCAAC CGCAGGATGA GCCAAAACTT
 401  TCTCTTGTTG ACACCATTCC ACAATCTTCC CTTTCTCTTC AAGCTCTAGA
 451  GGCAAGACAT GCGGTTCTAT GGCTAAGCCT TCTAAGGGAG GCCGCAAAAC
 501  CCATAAGCAG GACAACCCGG AGTTGAGAAT GCCGTGAGCA ATCTCGTCGA
 551  TCTGGTTTTG CTTCAAGAAA GCCAAAGTCC CAAAAGAGAT GTAAACAACG
 601  GAGGATGGTT CTCTCGAGTC AAGCCACTCT ATGCAGTCAC TATCTGGCTT
 651  GGAGATGTCT CCCTTGATGT CAGACCTTAT GGTTTTAGCC ATAGTAAAAA
 701  GCGGACCGAT GGGGTTGAAG TTGACTTGAG GGCAGAGCTG GGACATGTGG
 751  TCAATGGTAT CTTTTTCAAG TTCTTGAAAA GTTTCGATGA GAACAGAGAA
 801  AGGCTTGTGA AGTCGCTTGA TCTGCTCTAA AATGGTACCT CCTATAGAGG
 851  ACAGCGGAGA GGAAGGGTGA AGAAAGCTAG GGATCTCGTC ATGCTTCAAT
 901  GTTAATGGCT TGAAAGGGAC GTCAACGGTT ATTTCCGGCT CGGTTTCGGT
 951  CGGAAACTTA ACTAACTGGT GGTGGTAATA GTAATAAGCG GCGAGGCAAG
1001  CACAAGACTG GACCCAAAGA ACAGCTGATG GGATTTGAAG CTCCTCGGCT
1051  ATGTCACAAA CCCATGGAAC AAAGGCATTA TTTATGAGAC ATCTCACTGG
1101  TTGCTTCTCA TATTTCTTGA CAAGATTCTT GATCTCTCGT TTTCCGGAAA
1151  CTTCAAGTGA TTTTTGTAAC AAATCAAAGT CTTCTTTGTA GACAAATCCA
1201  TCCTCGAAGA ACTCGAACCG GAGAAAACCT AGCCCGACCG GTTTGAGCAC
1251  ACCGTCTTGA ATATTGTTGG CTTGACGCAT CTTCTTGCCC AATGGTTCCT
1301  CTGTGGTTAC AAAGGTGACG ATTAAGCCTT TAGAGGCAAT GATCTTTCCG
1351  AGACGAAGAA GAGGGCTTAT GTGACCTTGC CCTGGGAATG ATACGAGCAT
1401  CACATGAGGT AACGACGATT CCATCTCCAT
```

FIGURE 6A **A911 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGGGCAGTA GTGAGGGTCA AGAAACACAT GTCCTAATGG TAACACTACC
  51  ATTCCAAGGT CACATCAATC CAATGCTCAA ACTCGCAAAA CATCTCTCGT
 101  TATCATCAAA GAACCTACAC ATCAATCTCG CCACTATTGA GTCAGCCCGT
 151  GATCTCCTCT CCACCGTAGA AAAACCTCGT TATCCGGTGG ACCTCGTGTT
 201  CTTCTCCGAT GGTCTACCTA AAGAAGATCC AAAGGCCCCT GAAACTCTTT
 251  TGAAGTCATT GAATAAAGTC GGAGCCATGA ACTTGTCTAA AATCATCGAA
 301  GAAAAGAGAT ACTCTTGTAT CATCTCTTCG CCTTTTACTC CATGGGTTCC
 351  AGCTGTTGCA GCCTCTCATA ACATCTCTTG TGCAATACTT TGGATCCAAG
 401  CTTGTGGAGC TTACTCGGTT TATTACCGTT ACTACATGAA GACAAACTCT
 451  TTCCCTGATC TTGAAGATCT GAATCAAACG GTGGAGTTAC CAGCTTTACC
 501  ATTGTTGGAA GTTCGAGATC TTCCATCGTT TATGTTACCT TCTGGTGGTG
 551  CTCACTTCTA TAATCTAATG GCGGAATTTG CAGATTGTTT GAGGTATGTG
 601  AAATGGGTTT TGGTTAATTC ATTCTATGAA CTCGAATCAG AGATAATCGA
 651  ATCGATGGCT GATTTAAAAC CTGTAATTCC AATTGGTCCT CTGGTTTCTC
 701  CATTTCTGTT GGGCGATGGT GAGGAGGAAA CCCTAGACGG TAAAAACCTA
 751  GATTTTTGTA AATCTGATGA TTGTTGTATG GAGTGGCTTG ACAAGCAAGC
 801  TAGGTCTTCT GTTGTGTACA TATCTTTCGG AAGTATGCTC GAAACATTGG
 851  AGAATCAGGT CGAGACCATA GCGAAGGCGC TGAAGAACAG AGGACTTCCA
 901  TTTCTTTGGG TGATAAGGCC AAAGGAGAAA GCCCAAAACG TTGCTGTTTT
 951  GCAGGAGATG GTGAAAGAAG GACAAGGGGT TGTTCTCGAG TGGAGTCCAC
1001  AAGAGAAGAT TTTGAGCCAC GAGGCAATCT CTTGTTTTGT CACGCATTGC
1051  GGCTGGAACT CGACTATGGA GACGGTGGTG GCTGGTGTTC CTGTGGTAGC
1101  GTACCCTAGC TGGACGGATC AGCCCATTGA CGCGCGGTTG CTTGTTGATG
1151  TGTTTGGAAT CGGAGTAAGG ATGAGGAATG ACAGTGTCGA TGGCGAGCTT
1201  AAGGTCGAAG AAGTAGAAAG ATGCATTGAG GCCGTGACGG AGGGACCCGC
1251  TGCCGTGGAT ATAAGAAGGA GAGCGGCGGA GCTAAAGCGC GTGGCGAGAT
1301  TGGCGTTGGC ACCTGGTGGA TCTTCGACAC GGAATTTAGA CTTGTTCATT
1351  AGTGATATCA CAATCGCCTA ACTCTTTACT TCAACTAGTA CAAAATGTAT
1401  GAATACAAGG TTTGATATAA CCACTATCAA TTGTTAG
```

FIGURE 6B **A911 AMINO ACID SEQUENCE**

```
   1  MGSSEGQETH VLMVTLPFQG HINPMLKLAK HLSLSSKNLH INLATIESAR
  51  DLLSTVEKPR YPVDLVFFSD GLPKEDPKAP ETLLKSLNKV GAMNLSKIIE
 101  EKRYSCIISS PFTPWVPAVA ASHNISCAIL WIQACGAYSV YYRYYMKTNS
 151  FPDLEDLNQT VELPALPLLE VRDLPSFMLP SGGAHFYNLM AEFADCLRYV
 201  KWVLVNSFYE LESEIIESMA DLKPVIPIGP LVSPFLLGDG EEETLDGKNL
 251  DFCKSDDCCM EWLDKQARSS VVYISFGSML ETLENQVETI AKALKNRGLP
 301  FLWVIRPKEK AQNVAVLQEM VKEGQGVVLE WSPQEKILSH EAISCFVTHC
 351  GWNSTMETVV AGVPVVAYPS WTDQPIDARL LVDVFGIGVR MRNDSVDGEL
 401  KVEEVERCIE AVTEGPAAVD IRRRAAELKR VARLALAPGG SSTRNLDLFI
 451  SDITIA
```

FIGURE 6C **A911 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  CTAACAATTG ATAGTGGTTA TATCAAACCT TGTATTCATA CATTTTGTAC
  51  TAGTTGAAGT AAAGAGTTAG GCGATTGTGA TATCACTAAT GAACAAGTCT
 101  AAATTCCGTG TCGAAGATCC ACCAGGTGCC AACGCCAATC TCGCCACGCG
 151  CTTTAGCTCC GCCGCTCTCC TTCTTATATC CACGGCAGCG GGTCCCTCCG
 201  TCACGGCCTC AATGCATCTT TCTACTTCTT CGACCTTAAG CTCGCCATCG
 251  ACACTGTCAT TCCTCATCCT TACTCCGATT CCAAACACAT CAACAAGCAA
 301  CCGCGCGTCA ATGGGCTGAT CCGTCCAGCT AGGGTACGCT ACCACAGGAA
 351  CACCAGCCAC CACCGTCTCC ATAGTCGAGT TCCAGCCGCA ATGCGTGACA
 401  AAACAAGAGA TTGCCTCGTG GCTCAAAATC TTCTCTTGTG GACTCCACTC
 451  GAGAACAACC CCTTGTCCTT CTTTCACCAT CTCCTGCAAA ACAGCAACGT
 501  TTTGGGCTTT CTCCTTTGGC CTTATCACCC AAAGAAATGG AAGTCCTCTG
 551  TTCTTCAGCG CCTTCGCTAT GGTCTCGACC TGATTCTCCA ATGTTTCGAG
 601  CATACTTCCG AAAGATATGT ACACAACAGA AGACCTAGCT TGCTTGTCAA
 651  GCCACTCCAT ACAACAATCA TCAGATTTAC AAAAATCTAG GTTTTTACCG
 701  TCTAGGGTTT CCTCCTCACC ATCGCCCAAC AGAAATGGAG AAACCAGAGG
 751  ACCAATTGGA ATTACAGGTT TTAAATCAGC CATCGATTCG ATTATCTCTG
 801  ATTCGAGTTC ATAGAATGAA TTAACCAAAA CCCATTTCAC ATACCTCAAA
 851  CAATCTGCAA ATTCCGCCAT TAGATTATAG AAGTGAGCAC CACCAGAAGG
 901  TAACATAAAC GATGGAAGAT CTCGAACTTC CAACAATGGT AAAGCTGGTA
 951  ACTCCACCGT TTGATTCAGA TCTTCAAGAT CAGGGAAAGA GTTTGTCTTC
1001  ATGTAGTAAC GGTAATAAAC CGAGTAAGCT CCACAAGCTT GGATCCAAAG
1051  TATTGCACAA GAGATGTTAT GAGAGGCTGC AACAGCTGGA ACCCATGGAG
1101  TAAAAGGCGA AGAGATGATA CAAGAGTATC TCTTTTCTTC GATGATTTTA
1151  GACAAGTTCA TGGCTCCGAC TTTATTCAAT GACTTCAAAA GAGTTTCAGG
1201  GGCCTTTGGA TCTTCTTTAG GTAGACCATC GGAGAAGAAC ACGAGGTCCA
1251  CCGGATAACG AGGTTTTTCT ACGGTGGAGA GGAGATCACG GGCTGACTCA
1301  ATAGTGGCGA GATTGATGTG TAGGTTCTTT GATGATAACG AGAGATGTTT
1351  TGCGAGTTTG AGCATTGGAT TGATGTGACC TTGGAATGGT AGTGTTACCA
1401  TTAGGACATG TGTTTCTTGA CCCTCACTAC TGCCCAT
```

FIGURE 7A **A119 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGCATATCA CAAAACCACA CGCCGCCATG TTTTCCAGTC CCGGAATGGG
  51  CCATGTCATC CCGGTGATCG AGCTTGGAAA GCGTCTCTCC GCTAACAACG
 101  GCTTCCACGT CACCGTCTTC GTCCTCGAAA CCGACGCAGC CTCCGCTCAA
 151  TCCAAGTTCC TAAACTCAAC CGGCGTCGAC ATCGTCAAAC TTCCATCGCC
 201  GGACATTTAT GGTTTAGTGG ACCCCGACGA CCATGTAGTG ACCAAGATCG
 251  GAGTCATTAT GCGTGCAGCA GTTCCAGCCC TCCGATCCAA GATCGCTGCC
 301  ATGCATCAAA AGCCAACGGC TCTGATCGTT GACTTGTTTG GCACAGATGC
 351  GTTATGTCTC GCAAAGGAAT TTAACATGTT GAGTTATGTG TTTATCCCTA
 401  CCAACGCACG TTTTCTCGGA GTTTCGATTT ATTATCCAAA TTTGGACAAA
 451  GATATCAAGG AAGAGCACAC AGTGCAAAGA AACCCACTCG CTATACCGGG
 501  GTGTGAACCG GTTAGGTTCG AAGATACTCT GGATGCATAT CTGGTTCCCG
 551  ACGAACCGGT GTACCGGGAT TTTGTTCGTC ATGGTCTGGC TTACCCAAAA
 601  GCCGATGGAA TTTTGGTAAA TACATGGGAA GAGATGGAGC CCAAATCATT
 651  GAAGTCCCTT CTAAACCCAA AGCTCTTGGG CCGGGTTGCT CGTGTACCGG
 701  TCTATCCAAT CGGTCCCTTA TGCAGACCGA TACAATCATC CGAAACCGAT
 751  CACCCGGTTT TGGATTGGTT AAACGAACAA CCGAACGAGT CGGTTCTCTA
 801  TATCTCCTTC GGGAGTGGTG GTTGTCTATC GGCGAAACAG TTAACTGAAT
 851  TGGCGTGGGG ACTCGAGCAG AGCCAGCAAC GGTTCGTATG GGTGGTTCGA
 901  CCACCGGTCG ACGGTTCGTG TTGTAGCGAG TATGTCTCGG CTAACGGTGG
 951  TGGAACCGAA GACAACACGC CAGAGTATCT ACCGGAAGGG TTCGTGAGTC
1001  GTACTAGTGA TAGAGGTTTC GTGGTCCCCT CATGGGCCCC ACAAGCTGAA
1051  ATCCTGTCCC ATCGGGCCGT TGGTGGGTTT TTGACCCATT GCGGTTGGAG
1101  CTCGACGTTG GAAAGCGTCG TTGGCGGCGT TCCGATGATC GCATGGCCAC
1151  TTTTTGCCGA GCAGAATATG AATGCGGCGT TGCTCAGCGA CGAACTGGGA
1201  ATCGCAGTCA GATTGGATGA TCCAAAGGAG GATATTTCTA GGTGGAAGAT
1251  TGAGGCGTTG GTGAGGAAGG TTATGACTGA GAAGGAAGGT GAAGCGATGA
1301  GAAGGAAAGT GAAGAAGTTG AGAGACTCGG CGGAGATGTC ACTGAGCATT
1351  GACGGTGGTG GTTTGGCGCA CGAGTCGCTT TGCAGAGTCA CCAAGGAGTG
1401  TCAACGGTTT TTGGAACGTG TCGTGGACTT GTCACGTGGT GCTTAG
```

FIGURE 7B **A119 AMINO ACID SEQUENCE**

```
  1  MHITKPHAAM FSSPGMGHVI PVIELGKRLS ANNGFHVTVF VLETDAASAQ
 51  SKFLNSTGVD IVKLPSPDIY GLVDPDDHVV TKIGVIMRAA VPALRSKIAA
101  MHQKPTALIV DLFGTDALCL AKEFNMLSYV FIPTNARFLG VSIYYPNLDK
151  DIKEEHTVQR NPLAIPGCEP VRFEDTLDAY LVPDEPVYRD FVRHGLAYPK
201  ADGILVNTWE EMEPKSLKSL LNPKLLGRVA RVPVYPIGPL CRPIQSSETD
251  HPVLDWLNEQ PNESVLYISF GSGGCLSAKQ LTELAWGLEQ SQQRFVWVVR
301  PPVDGSCCSE YVSANGGGTE DNTPEYLPEG FVSRTSDRGF VVPSWAPQAE
351  ILSHRAVGGF LTHCGWSSTL ESVVGGVPMI AWPLFAEQNM NAALLSDELG
401  IAVRLDDPKE DISRWKIEAL VRKVMTEKEG EAMRRKVKKL RDSAEMSLSI
451  DGGGLAHESL CRVTKECQRF LERVVDLSRG A
```

FIGURE 7C **A119 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  CTAAGCACCA CGTGACAAGT CCACGACACG TTCCAAAAAC CGTTGACACT
  51  CCTTGGTGAC TCTGCAAAGC GACTCGTGCG CCAAACCACC ACCGTCAATG
 101  CTCAGTGACA TCTCCGCCGA GTCTCTCAAC TTCTTCACTT TCCTTCTCAT
 151  CGCTTCACCT TCCTTCTCAG TCATAACCTT CCTCACCAAC GCCTCAATCT
 201  TCCACCTAGA AATATCCTCC TTTGGATCAT CCAATCTGAC TGCGATTCCC
 251  AGTTCGTCGC TGAGCAACGC CGCATTCATA TTCTGCTCGG CAAAAAGTGG
 301  CCATGCGATC ATCGGAACGC CGCCAACGAC GCTTTCCAAC GTCGAGCTCC
 351  AACCGCAATG GGTCAAAAAC CCACCAACGG CCCGATGGGA CAGGATTTCA
 401  GCTTGTGGGG CCCATGAGGG GACCACGAAA CCTCTATCAC TAGTACGACT
 451  CACGAACCCT TCCGGTAGAT ACTCTGGCGT GTTGTCTTCG GTTCCACCAC
 501  CGTTAGCCGA GACATACTCG CTACAACACG AACCGTCGAC CGGTGGTCGA
 551  ACCACCCATA CGAACCGTTG CTGGCTCTGC TCGAGTCCCC ACGCCAATTC
 601  AGTTAACTGT TTCGCCGATA GACAACCACC ACTCCCGAAG GAGATATAGA
 651  GAACCGACTC GTTCGGTTGT TCGTTTAACC AATCCAAAAC CGGGTGATCG
 701  GTTTCGGATG ATTGTATCGG TCTGCATAAG GGACCGATTG GATAGACCGG
 751  TACACGAGCA ACCCGGCCCA AGAGCTTTGG GTTTAGAAGG GACTTCAATG
 801  ATTTGGGCTC CATCTCTTCC CATGTATTTA CCAAAATTCC ATCGGCTTTT
 851  GGGTAAGCCA GACCATGACG AACAAAATCC CGGTACACCG GTTCGTCGGG
 901  AACCAGATAT GCATCCAGAG TATCTTCGAA CCTAACCGGT TCACACCCCG
 951  GTATAGCGAG TGGGTTTCTT TGCACTGTGT GCTCTTCCTT GATATCTTTG
1001  TCCAAATTTG GATAATAAAT CGAAACTCCG AGAAAACGTG CGTTGGTAGG
1051  GATAAACACA TAACTCAACA TGTTAAATTC CTTTGCGAGA CATAACGCAT
1101  CTGTGCCAAA CAAGTCAACG ATCAGAGCCG TTGGCTTTTG ATGCATGGCA
1151  GCGATCTTGG ATCGGAGGGC TGGAACTGCT GCACGCATAA TGACTCCGAT
1201  CTTGGTCACT ACATGGTCGT CGGGGTCCAC TAAACCATAA ATGTCCGGCG
1251  ATGGAAGTTT GACGATGTCG ACGCCGGTTG AGTTTAGGAA CTTGGATTGA
1301  GCGGAGGCTG CGTCGGTTTC GAGGACGAAG ACGGTGACGT GGAAGCCGTT
1351  GTTAGCGGAG AGACGCTTTC CAAGCTCGAT CACCGGGATG ACATGGCCCA
1401  TTCCGGGACT GGAAAACATG GCGGCGTGTG GTTTTGTGAT ATGCAT
```

FIGURE 8A **A233 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGAGTAGTG ATCCTCATCG TAAGCTCCAT GTTGTGTTCT TCCCTTTCAT
  51  GGCTTATGGT CACATGATAC CAACTCTAGA CATGGCTAAG CTTTTCTCTA
 101  GCAGAGGAGC CAAATCTACA ATCCTCACCA CACCTCTCAA CTCCAAGATC
 151  TTCCAAAAAC CCATCGAAAG ATTCAAGAAC CTGAATCCGA GTTTCGAAAT
 201  CGACATCCAG ATCTTCGATT TCCCTTGCGT GGATCTCGGG TTACCAGAAG
 251  GATGCGAAAA CGTCGATTTC TTCACCTCAA ACAACAATGA TGATAGACAG
 301  TATCTGACCT TGAAGTTCTT TAAGTCGACA AGGTTTTTCA AAGATCAGCT
 351  TGAGAAGCTC CTCGAGACAA CGAGACCAGA CTGTCTTATC GCCGACATGT
 401  TCTTCCCCTG GGCTACGGAA GCTGCTGAGA AGTTCAATGT GCCAAGACTT
 451  GTGTTCCACG GTACTGGCTA CTTTTCTTTA TGCTCTGAAT ATTGCATCAG
 501  AGTGCATAAC CCACAAAACA TAGTAGCTTC AAGGTACGAG CCATTTGTGA
 551  TTCCTGATCT CCCGGGGAAC ATAGTGATAA CTCAAGAACA GATAGCAGAC
 601  CGTGACGAAG AAAGCGAGAT GGGGAAGTTT ATGATTGAGG TCAAAGAATC
 651  TGATGTGAAG AGCTCAGGTG TTATTGTAAA CAGCTTCTAC GAGCTTGAAC
 701  CTGATTACGC CGACTTTTAC AAGAGTGTTG TACTGAAGAG AGCGTGGCAT
 751  ATCGGTCCGC TTTCGGTTTA CAACAGAGGA TTTGAGGAGA AGGCTGAGAG
 801  AGGAAAGAAA GCAAGCATTA ATGAGGTTGA ATGCCTCAAA TGGCTTGACT
 851  CCAAGAAACC AGATTCAGTC ATTTACATTT CTTTTGGGAG CGTGGCTTGC
 901  TTCAAGAACG AGCAGCTATT CGAGATCGCT GCAGGATTAG AAACTTCTGG
 951  AGCAAATTTC ATCTGGGTTG TTAGGAAAAA CATAGGTATT GAAAAAGAAG
1001  AATGGTTACC AGAAGGGTTC GAAGAGAGGG TGAAAGGAAA AGGGATGATT
1051  ATAAGAGGAT GGGCACCACA GGTGCTCATA CTTGATCATC AAGCAACTTG
1101  TGGGTTTGTG ACCCATTGCG GCTGGAACTC GCTTCTGGAA GGAGTGGCTG
1151  CAGGGCTACC AATGGTGACA TGGCCTGTAG CAGCGGAGCA ATTCTACAAT
1201  GAGAAATTGG TTACGCAAGT GCTCAGAACA GGAGTGAGCG TGGGAGCGAA
1251  AAAGAATGTA AGAACTACGG GAGATTTCAT TAGCAGAGAG AAAGTGGTTA
1301  AAGCGGTGAG GGAGGTGTTG GTTGGGGAAG AGGCGGATGA GAGGCGGGAG
1351  AGGGCAAAGA AGTTGGCAGA GATGGCTAAA GCTGCCGTGG AAGGAGGGTC
1401  TTCTTTCAAC GATCTAAACA GCTTCATAGA AGAGTTTACC TCGTAA
```

FIGURE 8B **A233 AMINO ACID SEQUENCE**

```
   1  MSSDPHRKLH VVFFPFMAYG HMIPTLDMAK LFSSRGAKST ILTTPLNSKI
  51  FQKPIERFKN LNPSFEIDIQ IFDFPCVDLG LPEGCENVDF FTSNNNDDRQ
 101  YLTLKFFKST RFFKDQLEKL LETTRPDCLI ADMFFPWATE AAEKFNVPRL
 151  VFHGTGYFSL CSEYCIRVHN PQNIVASRYE PFVIPDLPGN IVITQEQIAD
 201  RDEESEMGKF MIEVKESDVK SSGVIVNSFY ELEPDYADFY KSVVLKRAWH
 251  IGPLSVYNRG FEEKAERGKK ASINEVECLK WLDSKKPDSV IYISFGSVAC
 301  FKNEQLFEIA AGLETSGANF IWVVRKNIGI EKEEWLPEGF EERVKGKGMI
 351  IRGWAPQVLI LDHQATCGFV THCGWNSLLE GVAAGLPMVT WPVAAEQFYN
 401  EKLVTQVLRT GVSVGAKKNV RTTGDFISRE KVVKAVREVL VGEEADERRE
 451  RAKKLAEMAK AAVEGGSSFN DLNSFIEEFT S
```

FIGURE 8C **A233 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  TTACGAGGTA AACTCTTCTA TGAAGCTGTT TAGATCGTTG AAAGAAGACC
  51  CTCCTTCCAC GGCAGCTTTA GCCATCTCTG CCAACTTCTT TGCCCTCTCC
 101  CGCCTCTCAT CCGCCTCTTC CCCAACCAAC ACCTCCCTCA CCGCTTTAAC
 151  CACTTTCTCT CTGCTAATGA AATCTCCCGT AGTTCTTACA TTCTTTTTCG
 201  CTCCCACGCT CACTCCTGTT CTGAGCACTT GCGTAACCAA TTTCTCATTG
 251  TAGAATTGCT CCGCTGCTAC AGGCCATGTC ACCATTGGTA GCCCTGCAGC
 301  CACTCCTTCC AGAAGCGAGT TCCAGCCGCA ATGGGTCACA AACCCACAAG
 351  TTGCTTGATG ATCAAGTATG AGCACCTGTG GTGCCCATCC TCTTATAATC
 401  ATCCCTTTTC CTTTCACCCT CTCTTCGAAC CCTTCTGGTA ACCATTCTTC
 451  TTTTTCAATA CCTATGTTTT TCCTAACAAC CCAGATGAAA TTTGCTCCAG
 501  AAGTTTCTAA TCCTGCAGCG ATCTCGAATA GCTGCTCGTT CTTGAAGCAA
 551  GCCACGCTCC CAAAAGAAAT GTAAATGACT GAATCTGGTT TCTTGGAGTC
 601  AAGCCATTTG AGGCATTCAA CCTCATTAAT GCTTGCTTTC TTTCCTCTCT
 651  CAGCCTTCTC CTCAAATCCT CTGTTGTAAA CCGAAAGCGG ACCGATATGC
 701  CACGCTCTCT TCAGTACAAC ACTCTTGTAA AAGTCGGCGT AATCAGGTTC
 751  AAGCTCGTAG AAGCTGTTTA CAATAACACC TGAGCTCTTC ACATCAGATT
 801  CTTTGACCTC AATCATAAAC TTCCCCATCT CGCTTTCTTC GTCACGGTCT
 851  GCTATCTGTT CTTGAGTTAT CACTATGTTC CCCGGGAGAT CAGGAATCAC
 901  AAATGGCTCG TACCTTGAAG CTACTATGTT TTGTGGGTTA TGCACTCTGA
 951  TGCAATATTC AGAGCATAAA GAAAAGTAGC CAGTACCGTG GAACACAAGT
1001  CTTGGCACAT TGAACTTCTC AGCAGCTTCC GTAGCCCAGG GGAAGAACAT
1051  GTCGGCGATA AGACAGTCTG GTCTCGTTGT CTCGAGGAGC TTCTCAAGCT
1101  GATCTTTGAA AAACCTTGTC GACTTAAAGA ACTTCAAGGT CAGATACTGT
1151  CTATCATCAT TGTTGTTTGA GGTGAAGAAA TCGACGTTTT CGCATCCTTC
1201  TGGTAACCCG AGATCCACGC AAGGGAAATC GAAGATCTGG ATGTCGATTT
1251  CGAAACTCGG ATTCAGGTTC TTGAATCTTT CGATGGGTTT TTGGAAGATC
1301  TTGGAGTTGA GAGGTGTGGT GAGGATTGTA GATTTGGCTC CTCTGCTAGA
1351  GAAAAGCTTA GCCATGTCTA GAGTTGGTAT CATGTGACCA TAAGCCATGA
1401  AAGGGAAGAA CACAACATGG AGCTTACGAT GAGGATCACT ACTCAT
```

FIGURE 9A **A407 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGCATATCA CAAAACCACA CGCCGCCATG TTTTCCAGTC CCGGAATGGG
  51  CCATGTCCTC CCGGTGATCG AGCTAGCTAA GCGTCTCTCC GCTAACCACG
 101  GCTTCCACGT CACCGTCTTC GTCCTTGAAA CTGACGCAGC CTCCGTTCAG
 151  TCCAAGCTCC TTAACTCAAC CGGTGTTGAC ATCGTCAACC TTCCATCGCC
 201  CGACATTTCT GGCTTGGTAG ACCCCAACGC CCATGTGGTG ACCAAGATCG
 251  GAGTCATTAT GCGTGAAGCT GTTCCAACCC TCCGATCCAA GATCGTTGCC
 301  ATGCATCAAA ACCCAACGGC TCTGATCATT GACTTGTTTG GCACAGATGC
 351  GTTATGTCTT GCAGCGGAGT TAAACATGTT GACTTATGTC TTTATCGCTT
 401  CCAACGCGCG TTATCTCGGA GTTTCGATAT ATTATCCAAC TTTGGACGAA
 451  GTTATCAAAG AAGAGCACAC AGTGCAACGA AAACCGCTCA CTATACCGGG
 501  GTGTGAACCG GTTAGATTTG AAGATATTAT GGATGCATAT CTGGTTCCGG
 551  ACGAACCGGT GTACCACGAT TTGGTTCGTC ACTGTCTGGC CTACCCAAAA
 601  GCGGATGGAA TCTTGGTGAA TACATGGGAA GAGATGGAGC CCAAATCATT
 651  AAAGTCCCTT CAAGACCCGA AACTTTTGGG CCGGGTCGCT CGTGTACCGG
 701  TTTATCCGGT TGGTCCGTTA TGCAGACCGA TACAATCATC CACGACCGAT
 751  CACCCGGTTT TTGATTGGTT AAACAAACAA CCAAACGAGT CGGTTCTCTA
 801  CATTTCCTTC GGGAGTGGTG GTTCTCTAAC GGCTCAACAG TTAACCGAAT
 851  TGGCGTGGGG GCTCGAGGAG AGCCAGCAAC GGTTTATATG GGTGGTTCGA
 901  CCGCCCGTTG ACGGCTCGTC TTGCAGTGAT TATTTCTCGG CTAAAGGCGG
 951  TGTAACCAAA GACAACACGC CAGAGTATCT ACCAGAAGGG TTCGTGACTC
1001  GTACTTGCGA TAGAGGTTTC ATGATCCCAT CATGGGCACC GCAAGCTGAA
1051  ATCCTAGCCC ATCAGGCCGT TGGTGGGTTT TTAACACATT GTGGTTGGAG
1101  CTCGACGTTG GAAAGCGTCC TTTGCGGCGT TCCAATGATA GCGTGGCCGC
1151  TTTTCGCCGA GCAGAATATG AACGCGGCGT TGCTTAGCGA TGAACTGGGA
1201  ATCTCTGTTA GAGTGGATGA TCCAAAGGAG GCGATTTCTA GGTCGAAGAT
1251  TGAGGCGATG GTGAGGAAGG TTATGGCTGA GGACGAAGGT GAAGAGATGA
1301  GAAGGAAAGT GAAGAAGTTG AGAGACACGG CGGAGATGTC ACTTAGTATT
1351  CACGGTGGTG GTTCGGCGCA TGAGTCGCTT TGCAGAGTCA CGAAGGAGTG
1401  TCAACGGTTT TTGGAATGTG TCGGGGACTT GGGACGTGGT GCTTAG
```

FIGURE 9B **A407 AMINO ACID SEQUENCE**

```
  1  MHITKPHAAM FSSPGMGHVL PVIELAKRLS ANHGFHVTVF VLETDAASVQ
 51  SKLLNSTGVD IVNLPSPDIS GLVDPNAHVV TKIGVIMREA VPTLRSKIVA
101  MHQNPTALII DLFGTDALCL AAELNMLTYV FIASNARYLG VSIYYPTLDE
151  VIKEEHTVQR KPLTIPGCEP VRFEDIMDAY LVPDEPVYHD LVRHCLAYPK
201  ADGILVNTWE EMEPKSLKSL QDPKLLGRVA RVPVYPVGPL CRPIQSSTTD
251  HPVFDWLNKQ PNESVLYISF GSGGSLTAQQ LTELAWGLEE SQQRFIWVVR
301  PPVDGSSCSD YFSAKGGVTK DNTPEYLPEG FVTRTCDRGF MIPSWAPQAE
351  ILAHQAVGGF LTHCGWSSTL ESVLCGVPMI AWPLFAEQNM NAALLSDELG
401  ISVRVDDPKE AISRSKIEAM VRKVMAEDEG EEMRRKVKKL RDTAEMSLSI
451  HGGGSAHESL CRVTKECQRF LECVGDLGRG A
```

FIGURE 9C **A407 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  CTAAGCACCA CGTCCCAAGT CCCCGACACA TTCCAAAAAC CGTTGACACT
  51  CCTTCGTGAC TCTGCAAAGC GACTCATGCG CCGAACCACC ACCGTGAATA
 101  CTAAGTGACA TCTCCGCCGT GTCTCTCAAC TTCTTCACTT TCCTTCTCAT
 151  CTCTTCACCT TCGTCCTCAG CCATAACCTT CCTCACCATC GCCTCAATCT
 201  TCGACCTAGA AATCGCCTCC TTTGGATCAT CCACTCTAAC AGAGATTCCC
 251  AGTTCATCGC TAAGCAACGC CGCGTTCATA TTCTGCTCGG CGAAAAGCGG
 301  CCACGCTATC ATTGGAACGC CGCAAAGGAC GCTTTCCAAC GTCGAGCTCC
 351  AACCACAATG TGTTAAAAAC CCACCAACGG CCTGATGGGC TAGGATTTCA
 401  GCTTGCGGTG CCCATGATGG GATCATGAAA CCTCTATCGC AAGTACGAGT
 451  CACGAACCCT TCTGGTAGAT ACTCTGGCGT GTTGTCTTTG GTTACACCGC
 501  CTTTAGCCGA GAAATAATCA CTGCAAGACG AGCCGTCAAC GGGCGGTCGA
 551  ACCACCCATA TAAACCGTTG CTGGCTCTCC TCGAGCCCCC ACGCCAATTC
 601  GGTTAACTGT TGAGCCGTTA GAGAACCACC ACTCCCGAAG GAAATGTAGA
 651  GAACCGACTC GTTTGGTTGT TTGTTTAACC AATCAAAAAC CGGGTGATCG
 701  GTCGTGGATG ATTGTATCGG TCTGCATAAC GGACCAACCG GATAAACCGG
 751  TACACGAGCG ACCCGGCCCA AAAGTTTCGG GTCTTGAAGG GACTTTAATG
 801  ATTTGGGCTC CATCTCTTCC CATGTATTCA CCAAGATTCC ATCCGCTTTT
 851  GGGTAGGCCA GACAGTGACG AACCAAATCG TGGTACACCG GTTCGTCCGG
 901  AACCAGATAT GCATCCATAA TATCTTCAAA TCTAACCGGT TCACACCCCG
 951  GTATAGTGAG CGGTTTTCGT TGCACTGTGT GCTCTTCTTT GATAACTTCG
1001  TCCAAAGTTG GATAATATAT CGAAACTCCG AGATAACGCG CGTTGGAAGC
1051  GATAAAGACA TAAGTCAACA TGTTTAACTC CGCTGCAAGA CATAACGCAT
1101  CTGTGCCAAA CAAGTCAATG ATCAGAGCCG TTGGGTTTTG ATGCATGGCA
1151  ACGATCTTGG ATCGGAGGGT TGGAACAGCT TCACGCATAA TGACTCCGAT
1201  CTTGGTCACC ACATGGGCGT TGGGGTCTAC CAAGCCAGAA ATGTCGGGCG
1251  ATGGAAGGTT GACGATGTCA ACACCGGTTG AGTTAAGGAG CTTGGACTGA
1301  ACGGAGGCTG CGTCAGTTTC AAGGACGAAG ACGGTGACGT GGAAGCCGTG
1351  GTTAGCGGAG AGACGCTTAG CTAGCTCGAT CACCGGGAGG ACATGGCCCA
1401  TTCCGGGACT GGAAAACATG GCGGCGTGTG GTTTTGTGAT ATGCAT
```

FIGURE 10A **A961 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGGGGAAGC AAGAAGATGC AGAGCTCGTC ATCATACCTT TCCCTTTCTC
  51  CGGACACATT CTCGCAACAA TCGAACTCGC CAAACGTCTC ATAAGTCAAG
 101  ACAATCCTCG GATCCACACC ATCACCATCC TCTATTGGGG ATTACCTTTT
 151  ATTCCTCAAG CTGACACAAT CGCTTTCCTC CGATCCCTAG TCAAAAATGA
 201  GCCTCGTATC CGTCTCGTTA CGTTGCCCGA AGTCCAAGAC CCTCCACCAA
 251  TGGAACTCTT TGTGGAATTT GCCGAATCTT ACATTCTTGA ATACGTCAAG
 301  AAAATGGTTC CCATCATCAG AGAAGCTCTC TCCACTCTCT TGTCTTCCCG
 351  CGATGAATCG GGTTCAGTTC GTGTGGCTGG ATTGGTTCTT GACTTCTTCT
 401  GCGTCCCTAT GATCGATGTA GGAAACGAGT TTAATCTCCC TTCTTACATT
 451  TTCTTGACGT GTAGCGCAGG GTTCTTGGGT ATGATGAAGT ATCTTCCAGA
 501  GAGACACCGC GAAATCAAAT CGGAATTCAA CCGGAGCTTC AACGAGGAGT
 551  TGAATCTCAT TCCTGGTTAT GTCAACTCTG TTCCTACTAA GGTTTTGCCG
 601  TCAGGTCTAT TCATGAAAGA GACCTACGAG CCTTGGGTCG AACTAGCAGA
 651  GAGGTTTCCT GAAGCTAAGG GTATTTTGGT TAATTCATAC ACAGCTCTCG
 701  AGCCAAACGG TTTTAAATAT TTCGATCGTT GTCCGGATAA CTACCCAACC
 751  ATTTACCCAA TCGGGCCGAT ATTATGCTCC AACGACCGTC CGAATTTGGA
 801  CTCATCGGAA CGAGATCGGA TCATAACTTG GCTAGATGAC CAACCCGAGT
 851  CATCGGTCGT GTTCCTCTGT TTCGGGAGCT TGAAGAATCT CAGCGCTACT
 901  CAGATCAACG AGATAGCTCA AGCCTTAGAG ATCGTTGACT GCAAATTCAT
 951  CTGGTCGTTT CGAACCAACC CGAAGGAGTA CGCGAGCCCT TACGAGGCTC
1001  TACCACACGG GTTCATGGAC CGGGTCATGG ATCAAGGCAT TGTTTGTGGT
1051  TGGGCTCCTC AAGTTGAAAT CCTAGCCCAT AAAGCTGTGG GAGGATTCGT
1101  ATCTCATTGT GGTTGGAACT CGATATTGGA GAGTTTGGGT TTCGGCGTTC
1151  CAATCGCCAC GTGGCCGATG TACGCGGAAC AACAACTAAA CGCGTTCACG
1201  ATGGTGAAGG AGCTTGGTTT AGCCTTGGAG ATGCGGTTGG ATTACGTGTC
1251  GGAAGATGGA GATATAGTGA AAGCTGATGA GATCGCAGGA ACCGTTAGAT
1301  CTTTAATGGA CGGTGTGGAT GTGCCGAAGA GTAAAGTGAA GGAGATTGCT
1351  GAGGCGGGAA AAGAAGCTGT GGACGGTGGA TCTTCGTTTC TTGCGGTTAA
1401  AAGATTCATC GGTGACTTGA TCGACGGCGT TTCTATAAGT AAGTAG
```

FIGURE 10B **A961 AMINO ACID SEQUENCE**

```
  1  MGKQEDAELV IIPFPFSGHI LATIELAKRL ISQDNPRIHT ITILYWGLPF
 51  IPQADTIAFL RSLVKNEPRI RLVTLPEVQD PPPMELFVEF AESYILEYVK
101  KMVPIIREAL STLLSSRDES GSVRVAGLVL DFFCVPMIDV GNEFNLPSYI
151  FLTCSAGFLG MMKYLPERHR EIKSEFNRSF NEELNLIPGY VNSVPTKVLP
201  SGLFMKETYE PWVELAERFP EAKGILVNSY TALEPNGFKY FDRCPDNYPT
251  IYPIGPILCS NDRPNLDSSE RDRIITWLDD QPESSVVFLC FGSLKNLSAT
301  QINEIAQALE IVDCKFIWSF RTNPKEYASP YEALPHGFMD RVMDQGIVCG
351  WAPQVEILAH KAVGGFVSHC GWNSILESLG FGVPIATWPM YAEQQLNAFT
401  MVKELGLALE MRLDYVSEDG DIVKADEIAG TVRSLMDGVD VPKSKVKEIA
451  EAGKEAVDGG SSFLAVKRFI GDLIDGVSIS K
```

FIGURE 10C **A961 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  CTACTTACTT ATAGAAACGC CGTCGATCAA GTCACCGATG AATCTTTTAA
  51  CCGCAAGAAA CGAAGATCCA CCGTCCACAG CTTCTTTTCC CGCCTCAGCA
 101  ATCTCCTTCA CTTTACTCTT CGGCACATCC ACACCGTCCA TTAAAGATCT
 151  AACGGTTCCT GCGATCTCAT CAGCTTTCAC TATATCTCCA TCTTCCGACA
 201  CGTAATCCAA CCGCATCTCC AAGGCTAAAC CAAGCTCCTT CACCATCGTG
 251  AACGCGTTTA GTTGTTGTTC CGCGTACATC GGCCACGTGG CGATTGGAAC
 301  GCCGAAACCC AAACTCTCCA ATATCGAGTT CCAACCACAA TGAGATACGA
 351  ATCCTCCCAC AGCTTTATGG GCTAGGATTT CAACTTGAGG AGCCCAACCA
 401  CAAACAATGC CTTGATCCAT GACCCGGTCC ATGAACCCGT GTGGTAGAGC
 451  CTCGTAAGGG CTCGCGTACT CCTTCGGGTT GGTTCGAAAC GACCAGATGA
 501  ATTTGCAGTC AACGATCTCT AAGGCTTGAG CTATCTCGTT GATCTGAGTA
 551  GCGCTGAGAT TCTTCAAGCT CCCGAAACAG AGGAACACGA CCGATGACTC
 601  GGGTTGGTCA TCTAGCCAAG TTATGATCCG ATCTCGTTCC GATGAGTCCA
 651  AATTCGGACG GTCGTTGGAG CATAATATCG GCCCGATTGG GTAAATGGTT
 701  GGGTAGTTAT CCGGACAACG ATCGAAATAT TTAAAACCGT TTGGCTCGAG
 751  AGCTGTGTAT GAATTAACCA AAATACCCTT AGCTTCAGGA AACCTCTCTG
 801  CTAGTTCGAC CCAAGGCTCG TAGGTCTCTT TCATGAATAG ACCTGACGGC
 851  AAAACCTTAG TAGGAACAGA GTTGACATAA CCAGGAATGA GATTCAACTC
 901  CTCGTTGAAG CTCCGGTTGA ATTCCGATTT GATTTCGCGG TGTCTCTCTG
 951  GAAGATACTT CATCATACCC AAGAACCCTG CGCTACACGT CAAGAAAATG
1001  TAAGAAGGGA GATTAAACTC GTTTCCTACA TCGATCATAG GGACGCAGAA
1051  GAAGTCAAGA ACCAATCCAG CCACACGAAC TGAACCCGAT TCATCGCGGG
1101  AAGACAAGAG AGTGGAGAGA GCTTCTCTGA TGATGGGAAC CATTTTCTTG
1151  ACGTATTCAA GAATGTAAGA TTCGGCAAAT TCCACAAAGA GTTCCATTGG
1201  TGGAGGGTCT TGGACTTCGG GCAACGTAAC GAGACGGATA CGAGGCTCAT
1251  TTTTGACTAG GGATCGGAGG AAAGCGATTG TGTCAGCTTG AGGAATAAAA
1301  GGTAATCCCC AATAGAGGAT GGTGATGGTG TGGATCCGAG GATTGTCTTG
1351  ACTTATGAGA CGTTTGGCGA GTTCGATTGT TGCGAGAATG TGTCCGGAGA
1401  AAGGGAAAGG TATGATGACG AGCTCTGCAT CTTCTTGCTT CCCCAT
```

FIGURE 11A **A962 SENSE NUCLEOTIDE SEQUENCE**

```
   1  ATGGCGAAGC AGCAAGAAGC AGAGCTCATC TTCATCCCAT TTCCAATCCC
  51  CGGACACATT CTCGCCACAA TCGAACTCGC GAAACGTCTC ATCAGTCACC
 101  AACCTAGTCG GATCCACACC ATCACCATCC TCCATTGGAG CTTACCTTTT
 151  CTTCCTCAAT CTGACACTAT CGCCTTCCTC AAATCCCTAA TCGAAACAGA
 201  GTCTCGTATC CGTCTCATTA CCTTACCCGA TGTCCAAAAC CCTCCACCAA
 251  TGGAGCTATT TGTGAAAGCT TCCGAATCTT ACATTCTTGA ATACGTCAAG
 301  AAAATGGTTC CTTTGGTCAG AAACGCTCTC TCCACTCTCT TGTCTTCTCG
 351  TGATGAATCG GATTCAGTTC ATGTCGCCGG ATTAGTTCTT GATTTCTTCT
 401  GTGTCCCTTT GATCGATGTC GGAAACGAGT TTAATCTCCC TTCTTACATC
 451  TTCTTGACGT GTAGCGCAAG TTTCTTGGGT ATGATGAAGT ATCTTCTGGA
 501  GAGAAACCGC GAAACCAAAC CGGAACTTAA CCGGAGCTCT GACGAGGAAA
 551  CAATATCAGT TCCTGGTTTT GTTAACTCCG TTCCGGTTAA AGTTTTGCCA
 601  CCGGGTTTGT TCACGACTGA GTCTTACGAA GCTTGGGTCG AAATGGCGGA
 651  AAGGTTCCCT GAAGCCAAGG GTATTTTGGT CAATTCATTT GAATCTCTAG
 701  AACGTAACGC TTTTGATTAT TTCGATCGTC GTCCGGATAA TTACCCACCC
 751  GTTTACCCAA TCGGGCCAAT TCTATGCTCC AACGATCGTC CGAATTTGGA
 801  TTTATCGGAA CGAGACCGGA TCTTGAAATG GCTCGATGAC CAACCCGAGT
 851  CATCTGTTGT GTTTCTCTGC TTCGGGAGCT TGAAGAGTCT CGCTGCGTCT
 901  CAGATTAAAG AGATCGCTCA AGCCTTAGAG CTCGTCGGAA TCAGATTCCT
 951  CTGGTCGATT CGAACGGACC CGAAGGAGTA CGCGAGCCCG AACGAGATTT
1001  TACCGGACGG GTTTATGAAC CGAGTCATGG GTTTGGGCCT TGTTTGTGGT
1051  TGGGCTCCTC AAGTTGAAAT TCTGGCCCAT AAAGCAATTG GAGGGTTCGT
1101  GTCACACTGC GGTTGGAACT CGATATTGGA GAGTTTGCGT TTCGGAGTTC
1151  CAATTGCCAC GTGGCCAATG TACGCGGAAC AACAACTAAA CGCGTTCACG
1201  ATTGTGAAGG AGCTTGGTTT GGCGTTGGAG ATGCGGTTGG ATTACGTGTC
1251  GGAATATGGA GAAATCGTGA AAGCTGATGA AATCGCAGGA GCCGTACGAT
1301  CTTTGATGGA CGGTGAGGAT GTGCCGAGGA GGAAACTGAA GGAGATTGCG
1351  GAGGCGGGAA AAGAGGCTGT GATGGACGGT GGATCTTCGT TTGTTGCGGT
1401  TAAAAGATTC ATAGATGGGC TTTGA
```

FIGURE 11B **A962 AMINO ACID SEQUENCE**

```
  1  MAKQQEAELI FIPFPIPGHI LATIELAKRL ISHQPSRIHT ITILHWSLPF
 51  LPQSDTIAFL KSLIETESRI RLITLPDVQN PPPMELFVKA SESYILEYVK
101  KMVPLVRNAL STLLSSRDES DSVHVAGLVL DFFCVPLIDV GNEFNLPSYI
151  FLTCSASFLG MMKYLLERNR ETKPELNRSS DEETISVPGF VNSVPVKVLP
201  PGLFTTESYE AWVEMAERFP EAKGILVNSF ESLERNAFDY FDRRPDNYPP
251  VYPIGPILCS NDRPNLDLSE RDRILKWLDD QPESSVVFLC FGSLKSLAAS
301  QIKEIAQALE LVGIRFLWSI RTDPKEYASP NEILPDGFMN RVMGLGLVCG
351  WAPQVEILAH KAIGGFVSHC GWNSILESLR FGVPIATWPM YAEQQLNAFT
401  IVKELGLALE MRLDYVSEYG EIVKADEIAG AVRSLMDGED VPRRKLKEIA
451  EAGKEAVMDG GSSFVAVKRF IDGL
```

FIGURE 11C **A962 ANTISENSE NUCLEOTIDE SEQUENCE**

```
   1  TCAAAGCCCA TCTATGAATC TTTTAACCGC AACAAACGAA GATCCACCGT
  51  CCATCACAGC CTCTTTTCCC GCCTCCGCAA TCTCCTTCAG TTTCCTCCTC
 101  GGCACATCCT CACCGTCCAT CAAAGATCGT ACGGCTCCTG CGATTTCATC
 151  AGCTTTCACG ATTTCTCCAT ATTCCGACAC GTAATCCAAC CGCATCTCCA
 201  ACGCCAAACC AAGCTCCTTC ACAATCGTGA ACGCGTTTAG TTGTTGTTCC
 251  GCGTACATTG GCCACGTGGC AATTGGAACT CCGAAACGCA AACTCTCCAA
 301  TATCGAGTTC CAACCGCAGT GTGACACGAA CCCTCCAATT GCTTTATGGG
 351  CCAGAATTTC AACTTGAGGA GCCCAACCAC AAACAAGGCC CAAACCCATG
 401  ACTCGGTTCA TAAACCCGTC CGGTAAAATC TCGTTCGGGC TCGCGTACTC
 451  CTTCGGGTCC GTTCGAATCG ACCAGAGGAA TCTGATTCCG ACGAGCTCTA
 501  AGGCTTGAGC GATCTCTTTA ATCTGAGACG CAGCGAGACT CTTCAAGCTC
 551  CCGAAGCAGA GAAACACAAC AGATGACTCG GGTTGGTCAT CGAGCCATTT
 601  CAAGATCCGG TCTCGTTCCG ATAAATCCAA ATTCGGACGA TCGTTGGAGC
 651  ATAGAATTGG CCCGATTGGG TAAACGGGTG GGTAATTATC CGGACGACGA
 701  TCGAAATAAT CAAAAGCGTT ACGTTCTAGA GATTCAAATG AATTGACCAA
 751  AATACCCTTG GCTTCAGGGA ACCTTTCCGC CATTTCGACC CAAGCTTCGT
 801  AAGACTCAGT CGTGAACAAA CCCGGTGGCA AAACTTTAAC CGGAACGGAG
 851  TTAACAAAAC CAGGAACTGA TATTGTTTCC TCGTCAGAGC TCCGGTTAAG
 901  TTCCGGTTTG GTTTCGCGGT TTCTCTCCAG AAGATACTTC ATCATACCCA
 951  AGAAACTTGC GCTACACGTC AAGAAGATGT AAGAAGGGAG ATTAAACTCG
1001  TTTCCGACAT CGATCAAAGG GACACAGAAG AAATCAAGAA CTAATCCGGC
1051  GACATGAACT GAATCCGATT CATCACGAGA AGACAAGAGA GTGGAGAGAG
1101  CGTTTCTGAC CAAAGGAACC ATTTTCTTGA CGTATTCAAG AATGTAAGAT
1151  TCGGAAGCTT TCACAAATAG CTCCATTGGT GGAGGGTTTT GGACATCGGG
1201  TAAGGTAATG AGACGGATAC GAGACTCTGT TTCGATTAGG GATTTGAGGA
1251  AGGCGATAGT GTCAGATTGA GGAAGAAAAG GTAAGCTCCA ATGGAGGATG
1301  GTGATGGTGT GGATCCGACT AGGTTGGTGA CTGATGAGAC GTTTCGCGAG
1351  TTCGATTGTG GCGAGAATGT GTCCGGGGAT TGGAAATGGG ATGAAGATGA
1401  GCTCTGCTTC TTGCTGCTTC GCCAT
```

UGT71B5 Figure 12

ATGAAGATTGAGCTTGTGTTCATACCTTTGCCGGGGATTGGTCATCTCAGGCCAACCGTGAAGCTAGCG
AAGCAACTCATAGGCAGCGAAAACCGTCTTTCGATCACCATAATCATCATCCCTTCAAGATTTGACGCC
GGTGATGCATCCGCCTGTATCGCATCTCTCACCACGTTGTCTCAAGATGATCGCCTCCATTACGAATCC
ATATCCGTCGCAAAACAACCACCAACCTCCGACCCGGATCCTGTTCCGGCTCAAGTGTACATAGAGAAA
CAAAAGACGAAAGTGAGAGATGCAGTCGCGGCGAGAATCGTCGATCCAACAAGAAAGCTCGCGGGATTC
GTGGTGGACATGTTCTGTTCCTCGATGATCGATGTAGCTAACGAGTTTGGAGTTCCGTGTTATATGGTA
TACACATCGAACGCTACGTTTTTAGGAACCATGCTTCACGTTCAACAAATGTACGATCAAAAGAAGTAT
GACGTCAGCGAGTTAGAAAACTCGGTCACCGAGTTGGAGTTTCCGTCTCTGACTCGTCCTTATCCAGTG
AAGTGTCTTCCTCATATCCTCACTTCAAAGGAGTGGTTACCTCTCTCTCTAGCTCAAGCTAGGTGTTTC
CGGAAGATGAAGGGTATTTTGGTAAATACAGTTGCTGAGCTTGAACCTCACGCTTTGAAAATGTTCAAT
ATTAATGGTGACGATCTTCCTCAAGTTTATCCTGTTGGACCAGTGTTGCATCTCGAAAACGGCAATGAC
GATGATGAGAAGCAATCGGAAATTTTGCGGTGGCTCGACGAGCAACCGTCTAAATCTGTTGTGTTTCTC
TGCTTTGGGAGCTTGGGAGGTTTCACTGAAGAACAAACAAGAGAAACCGCTGTGGCCCTAGATAGAAGC
GGTCAGCGGTTTCTTTGGTGTCTTCGTCACGCATCGCCAAATATAAAAACAGATCGTCCCAGAGATTAC
ACGAATCTTGAGGAGGTTTTACCGGAGGGGTTCTTGGAACGGACTTTGGATAGAGGGAAAGTGATTGGA
TGGGCACCACAAGTGGCGGTACTAGAGAAGCCGGCGATAGGAGGGTTTGTCACTCACTGCGGTTGGAAC
TCTATTTTAGAGAGCTTGTGGTTCGGTGTTCCAATGGTGACGTGGCCGCTATACGCGGAACAGAAGGTT
AACGCGTTTGAGATGGTTGAGGAGCTGGGTTTGGCGGTGGAGATACGGAAGTACTTAAAAGGAGATTTG
TTCGCCGGAGAGATGGAGACGGTTACCGCGGAGGATATAGAGAGAGCCATTAGGCGTGTGATGGAGCAA
GACAGTGACGTTAGGAACAACGTGAAAGAGATGGCGGAGAAGTGCCACTTCGCGTTAATGGACGGTGGA
TCTTCGAAGGCGGCTTTGGAAAAGTTTATTCAAGACGTGATAGAGAATATGGATTAA

UGT71C3 Figure 13

ATGAAAGCAGAAGCAGAGATCATCTTCGTTACATATCCATCCCCTGGTCATCTTCTTGTCTCCATTGAA
TTCGCTAAATCTCTCATCAAACGTGATGATCGCATCCACACCATCACCATCCTCTACTGGGCTTTACCT
CTCGCTCCTCAAGCCCACCTTTTCGCTAAGTCCCTCGTTGCTTCACAGCCTCGAATCCGTCTCCTTGCG
TTGCCTGATGTTCAAAACCCTCCACCATTGGAACTCTTCTTTAAAGCTCCCGAAGCTTATATTCTTGAG
TCCACCAAGAAAACAGTTCCTTTAGTCAGAGACGCTCTCTCCACTCTAGTTTCTTCACGTAAAGAATCC
GGTTCGGTTCGTGTAGTCGGTTTGGTTATCGATTTTTTTTGTGTTCCAATGATCGAAGTGGCAAACGAG
CTTAACCTTCCTTCTTACATCTTCCTAACGTGTAACGCTGGGTTTTTAAGTATGATGAAGTATCTCCCT
GAGAGACATCGCATAACCACTTCTGAGCTAGATTTAAGCTCCGGCAACGTAGAACATCCAATTCCTGGC
TACGTCTGCTCCGTGCCGACGAAGGTTTTGCCTCCAGGTCTATTCGTGAGAGAGTCCTACGAGGCTTGG
GTCGAGATTGCAGAGAAGTTCCCTGGAGCCAAGGGCATTTTGGTAAACTCAGTCACATGTCTTGAGCAG
AATGCATTTGATTACTTCGCTCGTCTTGATGAGAACTATCCTCCGGTTTACCCGGTCGGACCGGTTCTT
AGTTTGAAGGATCGTCCGTCTCCAAATCTGGACGCATCGGACCGGGATCGGATCATGAGATGGCTCGAG
GACCAGCCGGAGTCGTCAATTGTGTATATCTGCTTCGGAAGCCTCGGAATCATTGGCAAGCTGCAGATT
GAAGAGATAGCTGAAGCCTTGGAACTCACCGGCCACAGGTTTCTTTGGTCAATACGTACAAATCCGACG
GAGAAAGCGAGCCCGTACGATCTGTTGCCGGAGGGATTTCTCGATCGGACGGCCAGTAAGGGATTGGTG
TGTGATTGGGCCCCGCAAGTAGAAGTTCTGGCCCATAAAGCGCTCGGAGGATTCGTGTCTCACTGCGGT
TGGAACTCTGTACTGGAGAGCTTATGGTTCGGTGTTCCGATCGCCACGTGGCCAATGTACGCTGAGCAA
CAGTTAAACGCATTCTCGATGGTGAAGGAGTTAGGGTTAGCCGTGGAGCTGCGTTTAGACTACGTTTCG
GCGTACGGAGAGATAGTAAAAGCTGAGGAGATCGCGGGAGCCATACGATCATTGATGGACGGTGAGGAT
ACGCCGAGGAAGAGAGTGAAGGAGATGGCGGAAGCGGCGAGGAATGCTTTGATGGACGGAGGATCTTCG
TTTGTTGCGGTTAAACGATTTCTCGACGAGTTGATCGGCGGAGATGTTTAG

UGT71C5 Figure 14

ATGAAGACAGCAGAGCTCATATTCGTTCCTCTGCCGGAGACCGGCCATCTCTTGTCAACGATCGAGTTT
GGAAAGCGTCTACTCAATCTAGACCGTCGGATTTCTATGATTACAATCCTCTCCATGAATCTTCCTTAC
GCTCCTCACGCCGACGCTTCTCTTGCTTCGCTAACAGCCTCCGAGCCTGGTATCCGAATCATCAGTCTC
CCGGAGATCCACGATCCACCTCCGATCAAGCTTCTTGACACTTCCTCCGAGACTTACATCCTCGATTTC
ATCCATAAAAACATACCTTGTCTCAGAAAAACCATCCAAGATTTAGTCTCATCATCATCATCTTCCGGA
GGTGGTAGTAGTCATGTCGCCGGCTTGATTCTTGATTTCTTCTGCGTTGGTTTGATCGACATCGGCCGT
GAGGTAAACCTTCCTTCCTATATCTTCATGACTTCCAACTTTGGTTTCTTAGGGGTTCTACAGTATCTC
CCGGAACGACAACGTTTGACTCCGTCGGAGTTCGATGAGAGCTCCGGCGAGGAAGAGTTACATATTCCG
GCGTTTGTGAACCGTGTTCCCGCCAAGGTTCTGCCGCCAGGTGTGTTCGATAAACTCTCTTACGGGTCT
CTGGTCAAAATCGGCGAGCGATTACATGAAGCCAAGGGTATTTTGGTTAATTCATTTACCCAAGTGGAG
CCTTATGCTGCTGAACATTTTTCTCAAGGACGAGATTACCCTCACGTGTATCCTGTTGGGCCGGTTCTC
AACTTAACGGGCCGTACAAATCCGGGTCTAGCTTCGGCCCAATATAAAGAGATGATGAAGTGGCTTGAC
GAGCAACCAGACTCGTCGGTTTTGTTCCTGTGTTTCGGGAGCATGGGAGTCTTCCCTGCACCTCAGATC
ACAGAGATTGCTCACGCGCTCGAGCTTATCGGGTGCAGGTTCATCTGGGCGATCCGTACGAACATGGCG
GGAGATGGCGATCCTCAGGAGCCGCTTCCAGAAGGATTTGTCGATCGAACAATGGGCCGTGGAATTGTG
TGTAGTTGGGCTCCACAAGTGGATATCTTGGCCCACAAGGCAACAGGTGGATTCGTTTCTCACTGCGGG
TGGAATTCCGTCCAAGAGAGTCTATGGTACGGTGTACCTATTGCAACGTGGCCAATGTATGCGGAGCAA
CAACTGAACGCATTTGAGATGGTGAAGGAGTTGGGCTTAGCAGTGGAGATAAGGCTTGACTACGTGGCG
GATGGTGATAGGGTTACTTTGGAGATCGTGTCAGCCGATGAAATAGCCACAGCCGTCCGATCATTGATG
GATAGTGATAACCCCGTGAGAAAGAAGGTTATAGAAAAATCTTCAGTGGCGAGGAAAGCTGTTGGTGAT
GGTGGGTCTTCTACGGTGGCCACATGTAATTTTATCAAAGATATTCTTGGGGATCACTTTTGA

UGT71D1 Figure 15

ATGCGGAATGTAGAGCTCATCTTCATCCCCACACCAACCGTTGGTCATCTTGTTCCGTTTCTTGAATTT
GCTAGGCGTCTCATTGAGCAAGATGATAGGATCCGTATCACAATCCTCTTGATGAAACTACAAGGTCAG
TCTCATCTAGACACTTATGTTAAATCAATTGCCTCCTCTCAACCGTTTGTTAGATTCATTGATGTCCCT
GAGTTAGAGGAGAAACCTACACTTGGTAGTACACAATCTGTGGAAGCTTATGTGTATGATGTTATTGAG
AGAAATATCCCTCTTGTGAGGAATATAGTCATGGATATTTTAACTTCTCTTGCATTGGATGGAGTTAAG
GTCAAGGGATTAGTTGTTGACTTTTTCTGTCTCCCTATGATTGACGTTGCTAAAGATATAAGTCTCCCT
TTCTATGTGTTCTTGACTACAAATTCCGGGTTCTTAGCTATGATGCAGTATCTAGCAGATCGACATAGT
AGAGATACATCGGTTTTTGTAAGAAACTCGGAAGAAATGTTGTCGATACCTGGATTTGTAAACCCTGTC
CCAGCCAATGTTCTGCCGTCAGCTCTGTTTGTTGAAGATGGTTATGATGCTTACGTTAAGCTGGCCATA
TTGTTTACAAAGGCCAATGGAATCCTAGTGAATAGCTCCTTTGATATTGAGCCTTACTCTGTGAATCAT
TTTCTTCAAGAACAGAATTATCCTTCTGTTTATGCTGTTGGCCCCATATTTGACTTGAAAGCCCAGCCT
CATCCAGAGCAGGACCTAACCCGTCGTGACGAGTTGATGAAATGGCTTGATGATCAACCCGAGGCATCG
GTTGTATTCCTTTGTTTTGGGAGTATGGCAAGGTTAAGAGGTTCTCTAGTGAAGGAAATAGCTCATGGA
CTTGAGCTATGTCAATATAGATTCCTCTGGTCACTCCGTAAAGAAGAGGTGACAAAGGATGATTTGCCA
GAGGGGTTCCTTGACCGTGTCGATGGACGTGGAATGATATGTGGTTGGTCTCCTCAGGTAGAAATACTG
GCCCATAAGGCAGTGGGAGGCTTTGTTTCTCACTGTGGATGGAACTCAATAGTAGAGAGTTTGTGGTTT
GGCGTGCCAATTGTGACATGGCCAATGTATGCAGAGCAACAACTCAATGCGTTTCTGATGGTGAAGGAA
CTGAAGCTAGCTGTGGAGCTGAAGCTTGATTACAGGGTACATAGTGATGAGATAGTAAACGCAAACGAG
ATAGAGACCGCTATTCGTTATGTAATGGACACGGATAATAATGTTGTGAGGAAACGAGTGATGGATATC
TCGCAGATGATCCAGAGAGCTACGAAGAATGGTGGATCTTCGTTTGCCGCAATTGAGAAATTCATATAT
GACGTGATAGGAATTAAGCCCTAG

UGT73B1 Figure 16

ATGGGAACTCCTGTCGAAGTCTCTAAGCTCCATTTCTTGCTCTTCCCTTTCATGGCTCATGGCCATATG
ATACCAACTCTAGACATGGCTAAGCTCTTTGCCACCAAAGGAGCTAAATCCACTATCCTCACTACACCT
CTCAATGCCAAGCTCTTCTTCGAGAAACCCATCAAATCATTCAACCAAGACAACCCGGGACTCGAAGAC
ATCACCATCCAGATCCTTAATTTCCCTTGCACAGAGCTTGGTTTGCCTGATGGCTGTGAGAATACTGAT
TTCATCTTCTCCACACCTGACCTAAACGTAGGTGACTTGAGTCAAAAGTTTTTACTCGCAATGAAATAT
TTCGAAGAGCCACTAGAGGAGCTCCTCGTGACAATGAGACCAGACTGTCTTGTCGGTAACATGTTCTTC
CCTTGGTCCACTAAAGTTGCTGAGAAGTTCGGAGTACCGAGACTTGTGTTCCACGGCACAGGCTACTTC
TCTTTATGTGCTTCTCATTGCATAAGGCTCCCTAAGAATGTGGCAACAAGTTCTGAGCCCTTTGTGATT
CCTGATCTCCCGGGAGACATTTTGATTACAGAGGAACAGGTCATGGAGACAGAAGAAGAGTCTGTAATG
GGGAGGTTTATGAAGGCAATAAGAGACTCAGAGAGAGATAGCTTTGGCGTGTTGGTGAACAGCTTCTAC
GAGCTTGAACAGGCTTACTCAGATTATTTCAAGAGCTTTGTGGCGAAAAGAGCGTGGCATATCGGTCCG
CTTTCCTTAGGAAATAGAAAGTTCGAGGAGAAAGCAGAAAGAGGCAAAAAGGCAAGCATTGATGAGCAT
GAATGTTTGAAATGGCTCGACTCCAAGAAATGTGATTCAGTGATTTACATGGCCTTTGGAACCATGTCT
AGCTTTAAAAACGAGCAGCTGATAGAGATTGCAGCTGGTTTAGATATGTCAGGACATGATTTTGTCTGG
GTGGTTAACAGAAAAGGCAGCCAAGGTACCATAGACATCACTCTCTTTGCAGCAAAATCCTCTGTTTTT
GTTTTAGAGAAAAACCAATGATCTAATTAGGATTCTACTGTTTCAAACTCTAACTTTTGCGTTTGCATT
ACATATAAATAGTTGAGAAGGAAGATTGGTTACCAGAGGGGTTTGAAGAGAAGACCAAGGGAAAAGGAT
TGATAATCCGAGGGTGGGCGCCACAAGTGCTGATACTTGAGCACAAAGCAATTGGCGGATTTTTGACGC
ATTGTGGATGGAACTCGTTATTAGAAGGGGTGGCAGCGGGCCTGCCAATGGTGACATGGCCCGTGGGAG
CCGAGCAGTTCTACAACGAGAAATTGGTGACACAAGTGTTGAAAACAGGAGTGAGTGTGGGAGTGAAGA
AGATGATGCAAGTAGTTGGAGACTTCATTAGCAGAGAGAAAGTGGAGGGAGCGGTGAGGGAAGTGATGG
TTGGAGAAGAGAGGAGGAAACGGGCCAAGGAGTTAGCAGAAATGGCGAAAAATGCGGTGAAAGAAGGAG
GATCTTCAGATCTAGAGGTAGATAGGTTGATGGAAGAGCTTACGTTAGTTAAACTGCAAAAAGAGAAGG
TATAA

UGT73B2 Figure 17

ATGGGTAGTGATCATCATCATCGAAAGCTCCACGTTATGTTCTTCCCTTTCATGGCTTATGGTCACATG
ATACCAACTCTAGACATGGCTAAGCTTTTCTCTAGCAGAGGAGCCAAATCCACAATCCTCACCACATCT
CTCAACTCCAAGATCCTCCAAAAACCCATCGACACATTCAAGAATCTGAATCCGGGTCTCGAAATCGAC
ATCCAGATCTTCAATTTCCCTTGCGTGGAGCTGGGGTTACCAGAAGGATGTGAAAACGTTGATTTCTTC
ACTTCAAACAACAATGATGATAAAAACGAGATGATCGTGAAATTCTTTTTCTCGACAAGGTTTTTCAAA
GACCAGCTTGAGAAACTCCTCGGGACAACGAGACCAGACTGTCTTATCGCCGACATGTTCTTCCCCTGG
GCTACTGAAGCTGCTGGGAAGTTCAATGTGCCAAGACTTGTGTTCCACGGCACTGGCTACTTCTCTTTA
TGCGCTGGTTATTGCATCGGAGTGCATAAACCACAGAAGAGAGTGGCTTCAAGCTCTGAGCCATTTGTG
ATTCCCGAGCTCCCTGGGAACATTGTGATAACTGAAGAACAGATCATAGATGGCGATGGAGAATCCGAC
ATGGGAAAGTTTATGACTGAAGTTAGGGAATCGGAAGTGAAGAGCTCAGGAGTTGTTTTGAATAGTTTC
TACGAGCTAGAACATGATTACGCCGATTTTTACAAAAGTTGTGTACAAAAGAGAGCGTGGCATATCGGT
CCGCTATCGGTTTACAACAGGGGATTTGAGGAGAAGGCTGAGAGAGGAAAGAAAGCGAACATTGATGAG
GCTGAATGCCTCAAATGGCTTGACTCCAAGAAACCAAATTCAGTCATTTATGTTTCCTTTGGGAGCGTG
GCTTTCTTCAAGAATGAACAGTTATTCGAGATCGCTGCAGGGTTAGAAGCTTCCGGTACAAGTTTCATT
TGGGTTGTTAGGAAAACCAAAGGTATTGAAATTGACGTTTGAAGCCTATATTATATAGCTGTAATTTGG
GTAGCTTTGATTTTAATCTGACACAAGATTTGGTGTGAACAGATGATAGAGAAGAATGGTTACCAGAAG
GGTTCGAAGAGAGGGTGAAAGGGAAAGGTATGATAATAAGAGGATGGGCACCACAGGTGCTGATACTTG
ACCACCAAGCAACCGGTGGGTTTGTGACCCATTGCGGCTGGAACTCGCTTCTTGAAGGAGTGGCTGCAG
GGCTACCAATGGTGACATGGCCTGTAGGAGCGGAGCAATTCTACAATGAGAAATTGGTTACGCAAGTGC
TCAGAACAGGAGTGAGCGTGGGAGCGAGCAAGCATATGAAAGTTATGATGGGAGATTTCATTAGCAGAG
AGAAAGTGGATAAAGCGGTGAGGGAGGTTTTGGCTGGGGAAGCAGCAGAGGAGAGGCGGAGACGGGCAA
AGAAGCTAGCGGCGATGGCTAAAGCTGCCGTGGAAGAAGGAGGGTCTTCCTTCAACGATCTAAACAGCT
TCATGGAAGAGTTTAGTTCATAA

UGT73B4 Figure 18

ATGAACAGAGAGCAAATTCATATTTTGTTCTTCCCCTTCATGGCTCATGGCCACATGATTCCACTCTTA
GACATGGCCAAGCTTTTCGCTAGAAGAGGAGCCAAATCAACTCTCCTCACAACCCCAATAAATGCTAAG
ATCTTGGAGAAACCCATTGAAGCATTCAAAGTTCAAAATCCTGATCTCGAAATCGGAATCAAGATCCTC
AATTTCCCTTGTGTAGAGCTTGGATTGCCAGAAGGATGCGAGAACCGTGACTTCATTAACTCATACCAA
AAATCTGACTCATTTGACTTGTTCTTGAAGTTTCTTTTCTCTACCAAGTATATGAAACAGCAGTTGGAG
AGTTTCATTGAAACAACCAAACCGAGTGCTCTTGTAGCCGATATGTTCTTCCCTTGGGCAACAGAATCC
GCGGAGAAGATCGGTGTTCCAAGACTTGTGTTCCACGGCACATCATCCTTTGCCTTGTGTTGTTCGTAT
AACATGAGGATTCATAAGCCACACAAGAAAGTCGCTTCGAGTTCTACTCCATTTGTAATCCCTGGTCTC
CCTGGAGACATAGTTATTACAGAAGACCAAGCCAATGTCACCAACGAAGAAACTCCATTCGGAAAGTTT
TGGAAAGAAGTCAGGGAATCAGAGACCAGTAGCTTTGGTGTTTTGGTGAATAGCTTCTACGAGCTGGAA
TCATCTTATGCTGATTTTTACCGTAGTTTTGTGGCGAAAAAAGCGTGGCATATAGGTCCACTTTCACTA
TCCAACAGAGGGATTGCAGAGAAAGCCGGAAGAGGGAAAAAGGCAAACATTGATGAGCAAGAATGCCTC
AAATGGCTTGACTCTAAGACACCTGGCTCAGTAGTTTACTTGTCCTTTGGTAGCGGAACCGGCTTACCC
AACGAACAGCTGTTAGAGATTGCTTTCGGCCTTGAAGGCTCTGGACAAAATTTCATTTGGGTGGTTAGC
AAAAATGAAAACCAAGGTAATTTTTTTCCTCCTTAACCATTATTAATCAATGTAGTCTTTATTAGTATA
TTTCCAAAAATATTAACATTTGTGTATACATTTTCCTATTGCCAAATATGCTATGATGCCATAGCAATG
AGTAGATTGGTTTGTGTACTTTATATATTACTTTGTAGAACTTCTAACAATTATGACTTGGTGTTGGTG
TAGTTGGGACAGGTGAAAATGAAGATTGGTTGCCTAAAGGGTTTGAAGAGAGGAATAAAGGAAAAGGGC
TGATAATACGCGGATGGGCCCCGCAAGTGCTGATACTTGACCACAAAGCAATCGGAGGATTTGTGACGC
ATTGCGGATGGAACTCGACTTTGGAGGGCATTGCCGCAGGGCTGCCTATGGTGACTTGGCCGATGGGGG
CAGAACAGTTCTACAACGAGAAGTTATTGACAAAAGTGTTGAGAATAGGAGTGAACGTTGGAGCTACCG
AGTTGGTGAAAAAAGGAAAGTTGATTAGTAGAGCACAAGTGGAGAAGGCAGTAAGGGAAGTGATTGGTG
GTGAGAAGGCAGAGGAAAGGCGGCTAAGGGCTAAGGAGCTGGGCGAGATGGCTAAAGCCGCTGTGGAAG
AAGGAGGGTCTTCTTATAATGATGTGAACAAGTTTATGGAAGAGCTGAATGGTAGAAAGTAG

UGT73B5 Figure 19

ATGAACAGAGAAGTCTCTGAGAGAATTCATATTTTGTTCTTCCCCTTCATGGCTCAAGGCCACATGATT
CCAATTTTGGACATGGCCAAGCTTTTCTCGAGGAGAGGAGCCAAGTCAACCCTTCTCACAACCCCAATC
AACGCTAAGATCTTCGAGAAACCTATTGAAGCATTCAAAAATCAAAACCCTGATCTCGAAATCGGAATC
AAGATCTTCAATTTCCCTTGTGTAGAGCTTGGATTGCCTGAAGGATGCGAGAACGCTGACTTTATCAAC
TCATACCAAAAATCTGACTCAGGTGACTTGTTCTTGAAGTTTCTTTTCTCTACCAAGTATATGAAACAA
CAGTTGGAGAGTTTCATTGAAACAACCAAACCAAGTGCTCTTGTTGCCGATATGTTCTTCCCTTGGGCG
ACAGAATCTGCTGAGAAGCTCGGTGTACCAAGACTTGTGTTCCACGGTACATCTTTCTTTTCTTTGTGT
TGTTCGTATAACATGAGGATTCATAAGCCACACAAGAAAGTCGCTACGAGTTCTACTCCTTTTGTAATC
CCTGGTCTCCCAGGAGACATAGTTATTACAGAAGACCAAGCCAATGTTGCCAAAGAAGAAACGCCAATG
GGAAAGTTTATGAAAGAGGTTAGGGAATCAGAGACCAATAGCTTTGGTGTATTGGTTAATAGCTTCTAC
GAGCTGGAATCAGCTTATGCTGATTTTTATCGTAGTTTTGTGGCGAAAAGAGCTTGGCATATCGGTCCG
CTTTCGCTATCTAACAGAGAGTTAGGAGAGAAAGCCAGAAGAGGGAAAAAGGCTAACATTGATGAGCAA
GAATGCCTAAAATGGCTGGACTCTAAGACACCTGGTTCAGTAGTTTACTTGTCCTTTGGGAGCGGAACT
AATTTCACCAACGACCAGCTGTTAGAGATCGCTTTTGGTCTTGAAGGTTCTGGACAAAGTTTCATCTGG
GTGGTTAGGAAAAATGAAAACCAAGGTAAATTGTTTCTCCCCAGCCATTATTAACCAACATAGTAATGT
TAATATTTGTGTATATATTCGTATTGCCAAATATGCTCTGATACCATGGCAAGTAATAGATTGGCTCAT
GTATTTTATTTGTGATCATGTAGAATTTTCTTAACAGTTATGACTTGGTGTTGGTATGGTTGGGACAGG
TGACAATGAAGAGTGGTTGCCTGAAGGGTTTAAAGAGAGGACAACAGGGAAAGGGCTAATAATACCTGG
ATGGGCGCCGCAAGTGCTGATACTTGACCATAAAGCAATTGGAGGATTTGTGACTCATTGCGGATGGAA
CTCGGCTATAGAGGGCATTGCCGCGGGGCTGCCTATGGTAACATGGCCAATGGGGGCAGAACAGTTCTA
CAATGAGAAGCTATTGACAAAAGTGTTGAGAATAGGAGTGAACGTTGGAGCTACCGAGTTGGTGAAAAA
AGGAAAGTTGATTAGTAGAGCACAAGTGGAGAAGGCAGTAAGGGAAGTGATTGGTGGTGAGAAGGCAGA
GGAAAGGCGGCTATGGGCTAAGAAGCTGGGCGAGATGGCTAAAGCCGCTGTGGAAGAAGGAGGGTCCTC
TTATAATGATGTGAACAAGTTTATGGAAGAGCTGAATGGTAGAAAGTAG

UGT73C1 Figure 20

ATGGCATCGGAATTTCGTCCTCCTCTTCATTTTGTTCTCTTCCCTTTCATGGCTCAAGGCCACATGATC
CCAATGGTAGATATTGCAAGGCTCCTGGCTCAGCGCGGGGTGACTATAACCATTGTCACTACACCTCAA
AACGCAGGCCGGTTCAAGAACGTTCTTAGCCGGGCTATCCAATCCGGCTTGCCCATCAATCTCGTGCAA
GTAAAGTTTCCATCTCAAGAATCGGGTTCACCGGAAGGACAGGAGAATTTGGACTTGCTCGATTCATTG
GGGGCTTCATTAACCTTCTTCAAAGCATTTAGCCTGCTCGAGGAACCAGTCGAGAAGCTCTTGAAAGAG
ATTCAACCTAGGCCAAACTGCATAATCGCTGACATGTGTTTGCCTTATACAAACAGAATTGCCAAGAAT
CTTGGTATACCAAAAATCATCTTTCATGGCATGTGTTGCTTCAATCTTCTTTGTACGCACATAATGCAC
CAAAACCACGAGTTCTTGGAAACTATAGAGTCTGACAAGGAATACTTCCCCATTCCTAATTTCCCTGAC
AGAGTTGAGTTCACAAAATCTCAGCTTCCAATGGTATTAGTTGCTGGAGATTGGAAAGACTTCCTTGAC
GGAATGACAGAAGGGGATAACACTTCTTATGGTGTGATTGTTAACACGTTTGAAGAGCTCGAGCCAGCT
TATGTTAGAGACTACAAGAAGGTTAAAGCGGGTAAGATATGGAGCATCGGACCGGTTTCCTTGTGCAAC
AAGTTAGGAGAAGACCAAGCTGAGAGGGGAAACAAGGCGGACATTGATCAAGACGAGTGTATTAAATGG
CTTGATTCTAAAGAAGAAGGGTCGGTGCTATATGTTTGCCTTGGAAGTATATGCAATCTTCCTCTGTCT
CAGCTCAAAGAGCTCGGCTTAGGCCTCGAGGAATCCCAAAGACCTTTCATTTGGGTCATAAGAGGTTGG
GAGAAGTATAACGAGTTACTTGAATGGATCTCAGAGAGCGGTTATAAGGAAAGAATCAAAGAAAGAGGC
CTTCTCATAACAGGATGGTCGCCTCAAATGCTTATCCTTACACATCCTGCCGTTGGAGGATTCTTGACA
CATTGTGGATGGAACTCTACTCTTGAAGGAATCACTTCAGGCGTTCCATTACTCACGTGGCCACTGTTT
GGAGACCAATTCTGCAATGAGAAATTGGCGGTGCAGATACTAAAAGCCGGTGTGAGAGCTGGGGTTGAA
GAGTCCATGAGATGGGGAGAAGAGGAGAAAATAGGAGTACTGGTGGATAAAGAAGGAGTAAAGAAGGCA
GTGGAGGAATTGATGGGTGATAGTAATGATGCTAAGGAGAGAAGAAAAAGAGTGAAAGAGCTTGGAGAA
TTAGCTCACAAGGCTGTGGAAGAAGGAGGCTCTTCTCATTCCAACATCACATTCTTGCTACAAGACATA
ATGCAATTAGAACAACCCAAGAAATGA

UGT731C Figure 21

ATGGCTACGGAAAAAACCCACCAATTTCATCCTTCTCTTCACTTTGTCCTCTTCCCTTTCATGGCTCAA
GGCCACATGATTCCCATGATTGATATTGCAAGACTCTTGGCTCAGCGTGGTGTGACCATAACAATTGTC
ACGACACCTCACAACGCAGCAAGGTTTAAGAATGTCCTAAACCGAGCGATCGAGTCTGGCTTGGCCATC
AACATACTGCATGTGAAGTTTCCATATCAAGAGTTTGGTTTGCCAGAAGGAAAAGAGAATATAGATTCG
TTAGACTCAACGGAGTTGATGGTACCTTTCTTCAAAGCGGTGAACTTGCTTGAAGATCCGGTCATGAAG
CTCATGGAAGAGATGAAACCTAGACCTAGCTGTCTAATTTCTGATTGGTGTTTGCCTTATACAAGCATA
ATCGCCAAGAACTTCAATATACCAAAGATAGTTTTCCACGGCATGGGTTGCTTTAATCTTTTGTGTATG
CATGTTCTACGCAGAAACTTAGAGATCCTAGAGAATGTAAAGTCGGATGAAGAGTATTTCTTGGTTCCT
AGTTTTCCTGATAGAGTTGAATTTACAAAGCTTCAACTTCCTGTGAAAGCAAATGCAAGTGGAGATTGG
AAAGAGATAATGGATGAAATGGTAAAAGCAGAATACACATCCTATGGTGTGATCGTCAACACATTTCAG
GAGTTGGAGCCACCTTATGTCAAAGACTACAAAGAGGCAATGGATGGAAAAGTATGGTCCATTGGACCC
GTTTCCTTGTGTAACAAGGCAGGTGCAGACAAAGCTGAGAGGGGAAGCAAGGCCGCCATTGATCAAGAT
GAGTGTCTTCAATGGCTTGATTCTAAAGAAGAAGGTTCGGTGCTCTATGTTTGCCTTGGAAGTATATGT
AATCTTCCTTTGTCTCAGCTCAAGGAGCTGGGGCTAGGCCTTGAGGAATCTCGAAGATCTTTTATTTGG
GTCATAAGAGGTTCGGAAAAGTATAAAGAACTATTTGAGTGGATGTTGGAGAGCGGTTTTGAAGAAAGA
ATCAAAGAGAGAGGACTTCTCATTAAAGGGTGGGCACCTCAAGTCCTTATCCTTTCACATCCTTCCGTT
GGAGGATTCCTGACACACTGTGGATGGAACTCGACTCTCGAAGGAATCACCTCAGGCATTCCACTGATC
ACTTGGCCGCTGTTTGGAGACCAATTCTGCAACCAAAAACTGGTCGTTCAAGTACTAAAAGCCGGTGTA
AGTGCCGGGGTTGAAGAAGTCATGAAATGGGGAGAAGAAGATAAAATAGGAGTGTTAGTGGATAAAGAA
GGAGTGAAAAAGGCTGTGGAAGAATTGATGGGTGATAGTGATGATGCAAAAGAGAGGAGAAGAAGAGTC
AAAGAGCTTGGAGAATTAGCTCACAAAGCTGTGGAAAAAGGAGGCTCTTCTCATTCTAACATCACACTC
TTGCTACAAGACATAATGCAACTAGCACAATTCAAGAATTGA

UGT73C5 Figure 22

ATGGTTTCCGAAACAACCAAATCTTCTCCACTTCACTTTGTTCTCTTCCCTTTCATGGCTCAAGGCCAC
ATGATTCCCATGGTTGATATTGCAAGGCTCTTGGCTCAGCGTGGTGTGATCATAACAATTGTCACGACG
CCTCACAATGCAGCGAGGTTCAAGAATGTCCTAAACCGTGCCATTGAGTCTGGCTTGCCCATCAACTTA
GTGCAAGTCAAGTTTCCATATCTAGAAGCTGGTTTGCAAGAAGGACAAGAGAATATCGATTCTCTTGAC
ACAATGGAGCGGATGATACCTTTCTTTAAAGCGGTTAACTTTCTCGAAGAACCAGTCCAGAAGCTCATT
GAAGAGATGAACCCTCGACCAAGCTGTCTAATTTCTGATTTTTGTTTGCCTTATACAAGCAAAATCGCC
AAGAAGTTCAATATCCCAAAGATCCTCTTCCATGGCATGGGTTGCTTTTGTCTTCTGTGTATGCATGTT
TTACGCAAGAACCGTGAGATCTTGGACAATTTAAAGTCAGATAAGGAGCTTTTCACTGTTCCTGATTTT
CCTGATAGAGTTGAATTCACAAGAACGCAAGTTCCGGTAGAAACATATGTTCCAGCTGGAGACTGGAAA
GATATCTTTGATGGTATGGTAGAAGCGAATGAGACATCTTATGGTGTGATCGTCAACTCATTTCAAGAG
CTCGAGCCTGCTTATGCCAAAGACTACAAGGAGGTAAGGTCCGGTAAAGCATGGACCATTGGACCCGTT
TCCTTGTGCAACAAGGTAGGAGCCGACAAAGCAGAGAGGGGAAACAAATCAGACATTGATCAAGATGAG
TGCCTTAAATGGCTCGATTCTAAGAAACATGGCTCGGTGCTTTACGTTTGTCTTGGAAGTATCTGTAAT
CTTCCTTTGTCTCAACTCAAGGAGCTGGGACTAGGCCTAGAGGAATCCCAAAGACCTTTCATTTGGGTC
ATAAGAGGTTGGGAGAAGTACAAAGAGTTAGTTGAGTGGTTCTCGGAAAGCGGCTTTGAAGATAGAATC
CAAGATAGAGGACTTCTCATCAAAGGATGGTCCCCTCAAATGCTTATCCTTTCACATCCATCAGTTGGA
GGGTTCCTAACACACTGTGGTTGGAACTCGACTCTTGAGGGGATAACTGCTGGTCTACCGCTACTTACA
TGGCCGCTATTCGCAGACCAATTCTGCAATGAGAAATTGGTCGTTGAGGTACTAAAAGCCGGTGTAAGA
TCCGGGGTTGAACAGCCTATGAAATGGGGAGAAGAGGAGAAAATAGGAGTGTTGGTGGATAAAGAAGGA
GTGAAGAAGGCAGTGGAAGAATTAATGGGTGAGAGTGATGATGCAAAAGAGAGAAGAAGAAGAGCCAAA
GAGCTTGGAGATTCAGCTCACAAGGCTGTGGAAGAAGGAGGCTCTTCTCATTCTAACATCTCTTTCTTG
CTACAAGACATAATGGAACTGGCAGAACCCAATAATTGA

UGT73C6 Figure 23

ATGGCTTTCGAAAAAAACAACGAACCTTTTCCTCTTCACTTTGTTCTCTTCCCTTTCATGGCTCAAGGC
CACATGATTCCCATGGTTGATATTGCAAGGCTCTTGGCTCAGCGAGGTGTGCTTATAACAATTGTCACG
ACGCCTCACAATGCAGCAAGGTTCAAGAATGTCCTAAACCGTGCCATTGAGTCTGGTTTGCCCATCAAC
CTAGTGCAAGTCAAGTTTCCATATCAAGAAGCTGGTCTGCAAGAAGGACAAGAAAATATGGATTTGCTT
ACCACGATGGAGCAGATAACATCTTTCTTTAAAGCGGTTAACTTACTCAAAGAACCAGTCCAGAACCTT
ATTGAAGAGATGAGCCCGCGACCAAGCTGTCTAATCTCTGATATGTGTTTGTCGTATACAAGCGAAATC
GCCAAGAAGTTCAAAATACCAAAGATCCTCTTCCATGGCATGGGTTGCTTTTGTCTTCTGTGTGTTAAC
GTTCTGCGCAAGAACCGTGAGATCTTGGACAATTTAAAGTCTGATAAGGAGTACTTCATTGTTCCTTAT
TTTCCTGATAGAGTTGAATTCACAAGACCTCAAGTTCCGGTGGAAACATATGTTCCTGCAGGCTGGAAA
GAGATCTTGGAGGATATGGTAGAAGCGGATAAGACATCTTATGGTGTTATAGTCAACTCATTTCAAGAG
CTCGAACCTGCGTATGCCAAAGACTTCAAGGAGGCAAGGTCTGGTAAAGCATGGACCATTGGACCTGTT
TCCTTGTGCAACAAGGTAGGAGTAGACAAAGCAGAGAGGGGAAACAAATCAGATATTGATCAAGATGAG
TGCCTTGAATGGCTCGATTCTAAGGAACCGGGATCTGTGCTCTACGTTTGCCTTGGAAGTATTTGTAAT
CTTCCTCTGTCTCAGCTCCTTGAGCTGGGACTAGGCCTAGAGGAATCCCAAAGACCTTTCATCTGGGTC
ATAAGAGGTTGGGAGAAATACAAAGAGTTAGTTGAGTGGTTCTCGGAAAGCGGCTTTGAAGATAGAATC
CAAGATAGAGGACTTCTCATCAAAGGATGGTCCCCTCAAATGCTTATCCTTTCACATCCTTCTGTTGGA
GGGTTCTTAACGCACTGCGGATGGAACTCGACTCTTGAGGGGATAACTGCTGGTCTACCAATGCTTACA
TGGCCACTATTTGCAGACCAATTCTGCAACGAGAAACTGGTCGTACAAATACTAAAAGTCGGTGTAAGT
GCCGAGGTTAAAGAGGTCATGAAATGGGGAGAAGAAGAGAAGATAGGAGTGTTGGTGGATAAAGAAGGA
GTGAAGAAGGCAGTGGAAGAACTAATGGGTGAGAGTGATGATGCAAAAGAGAGAAGAAGAAGAGCCAAA
GAGCTTGGAGAATCAGCTCACAAGGCTGTGGAAGAAGGAGGCTCCTCTCATTCTAATATCACTTTCTTG
CTACAAGACATAATGCAACTAGCACAGTCCAATAATTGA

UGT73C7 Figure 24

ATGTGTTCTCATGATCCTCTTCACTTCGTCGTAATACCCTTTATGGCCCAAGGCCATATGATCCCATTG
GTCGACATCTCTAGGCTCTTGTCCCAGCGCCAAGGCGTGACTGTCTGCATCATCACAACTACTCAAAAT
GTAGCCAAGATCAAGACTTCACTCTCATTTTCCTCTTTGTTTGCGACTATCAACATCGTTGAAGTTAAG
TTTCTGTCTCAACAAACGGGTTTGCCAGAAGGGTGCGAGAGTTTAGATATGTTGGCTTCAATGGGCGAT
ATGGTGAAGTTCTTTGATGCTGCCAACTCACTTGAGGAGCAAGTTGAGAAAGCTATGGAAGAGATGGTT
CAGCCGCGGCCAAGCTGCATCATTGGAGACATGAGCCTTCCTTTCACTTCAAGACTTGCCAAGAAATTC
AAGATCCCCAAACTTATCTTCCATGGGTTTTCTTGTTTCAGCCTCATGTCTATACAAGTGGTTCGAGAA
AGCGGGATCTTGAAAATGATAGAATCAAACGACGAGTATTTTGATTTGCCCGGCTTGCCTGACAAAGTT
GAGTTCACGAAACCTCAGGTCTCTGTGTTGCAACCTGTTGAAGGAAATATGAAAGAGAGTACGGCCAAG
ATTATTGAAGCTGATAATGACTCTTATGGTGTTATTGTGAACACTTTTGAAGAGTTAGAGGTTGATTAT
GCAAGAGAATATAGGAAAGCAAGGGCTGGAAAAGTTTGGTGCGTTGGACCTGTTTCCTTGTGCAATAGG
TTAGGGTTAGACAAAGCTAAAAGAGGAGATAAGGCTTCTATTGGTCAAGACCAATGTCTTCAATGGCTT
GACTCTCAAGAAACTGGTTCAGTGCTCTACGTTTGCCTTGGAAGTCTATGTAATCTTCCCTTGGCTCAG
CTCAAAGAGCTGGGACTAGGCCTTGAGGCATCTAATAAACCTTTCATATGGGTTATAAGAGAATGGGGA
AAATATGGAGATTTAGCAAATTGGATGCAACAAAGCGGATTTGAAGAGCGGATCAAAGATAGAGGACTG
GTGATCAAAGGTTGGGCGCCGCAAGTTTTCATCCTCTCACACGCATCCATTGGAGGGTTTTTGACTCAC
TGTGGATGGAACTCGACACTAGAAGGAATTACTGCAGGAGTTCCATTATTGACATGGCCTTTGTTTGCT
GAACAATTCTTGAATGAGAAGTTAGTTGTGCAGATACTAAAAGCAGGGTTAAAGATAGGAGTAGAGAAA
TTGATGAAATATGGAAAAGAAGAGGAGATAGGAGCGATGGTGAGCAGAGAATGTGTGAGAAAAGCTGTG
GATGAGCTAATGGGTGATAGTGAAGAAGCAGAAGAGAGAAGAAGAAAAGTTACAGAACTTAGTGACTTG
GCAAATAAGGCTTTGGAAAAAGGAGGATCTTCAGATTCTAATATCACATTGCTCATTCAAGATATTATG
GAGCAATCACAAAATCAATTTTAA

UGT74F2 Figure 25

ATGGAGCATAAGAGAGGACATGTATTAGCAGTGCCGTACCCAACGCAAGGACACATCACACCATTCCGC
CAATTCTGCAAACGACTTCACTTCAAAGGTCTCAAAACCACTCTCGCTCTCACCACTTTCGTCTTCAAC
TCCATCAATCCTGACCTATCCGGTCCAATCTCCATAGCCACCATCTCCGATGGCTATGACCATGGGGGT
TTCGAGACAGCTGACTCCATCGACGACTACCTCAAAGACTTTAAAACTTCCGGCTCGAAAACCATTGCA
GACATCATCCAAAAACACCAGACTAGTGATAACCCCATCACTTGTATCGTCTATGATGCTTTCCTGCCT
TGGGCACTTGACGTTGCTAGAGAGTTTGGTTTAGTTGCGACTCCTTTCTTTACGCAGCCTTGTGCTGTT
AACTATGTTTATTATCTTTCTTACATAAACAATGGAAGCTTGCAACTTCCCATTGAGGAATTGCCTTTT
CTTGAGCTCCAAGATTTGCCTTCTTTCTTCTCTGTTTCTGGCTCTTATCCTGCTTACTTTGAGATGGTG
CTTCAACAGTTCATAAATTTCGAAAAAGCTGATTTCGTTCTCGTTAATAGCTTCCAAGAGTTGGAACTG
CATGTTAGATCTCTCTCTATCTCTTTCTTACAATTCTTAAACCATCTCTTGTTCTTGTGCATGTACTAA
CTGCTCTTTTTTTGTTTACAGGAGAATGAATTGTGGTCGAAAGCTTGTCCTGTGTTGACAATTGGTCCA
ACTATTCCATCAATTTACTTAGACCAACGTATCAAATCAGACACCGGCTATGATCTTAATCTCTTTGAA
TCGAAAGATGATTCCTTCTGCATTAACTGGCTCGACACAAGGCCACAAGGGTCGGTGGTGTACGTAGCA
TTCGGAAGCATGGCTCAGCTGACTAATGTGCAGATGGAGGAGCTTGCTTCAGCAGTAAGCAACTTCAGC
TTCCTGTGGGTGGTCAGATCTTCAGAGGAGGAAAAACTCCCATCAGGGTTTCTTGAGACAGTGAATAAA
GAAAAGAGCTTGGTCTTGAAATGGAGTCCTCAGCTTCAAGTTCTGTCAAACAAAGCCATCGGTTGTTTC
TTGACTCACTGTGGCTGGAACTCAACCATGGAGGCTTTGACCTTCGGGGTTCCCATGGTGGCAATGCCC
CAATGGACTGATCAACCGATGAACGCAAAGTACATACAAGATGTGTGGAAGGCTGGAGTTCGTGTGAAG
ACAGAGAAGGAGAGTGGGATTGCCAAGAGAGAGGAGATTGAGTTTAGCATTAAGGAAGTGATGGAAGGA
GAGAGGAGCAAAGAGATGAAGAAGAACGTGAAGAAATGGAGAGACTTGGCTGTCAAGTCACTCAATGAA
GGAGGTTCTACGGATACTAACATTGATACATTTGTATCAAGGGTTCAGAGCAAATAG

UGT76E1 Figure 26

ATGGAAGAACTAGGAGTGAAGAGAAGGATAGTATTGGTTCCAGTTCCAGCACAAGGTCATGTAACTCCG
ATTATGCAACTCGGGAAGGCTCTTTACTCCAAGGGCTTCTCCATCACTGTTGTTCTCACACAGTATAAT
CGAGTTAGCTCATCCAAGGACTTCTCTGATTTTCATTTCCTCACCATCCCAGGCAGCTTGACCGAGTCT
GATCTCAAAAACCTTGGACCATTCAAGTTTCTCTTCAAGCTCAATCAAATTTGCGAGGCAAGCTTCAAG
CAATGTATTGGTCAACTATTGCAGGAGCAAGGTAATGATATCGCTTGTGTCGTCTACGATGAGTACATG
TACTTCTCCCAAGCTGCAGTTAAAGAGTTTCAACTTCCTAGCGTCCTCTTCAGCACGACAAGTGCTACT
GCCTTTGTCTGTCGCTCTGTTTTGTCTAGAGTCAACGCAGAGTCATTCTTGCTTGACATGAAAGGTACT
CAAGATTTTTTAGCTTGTTAACTCAAACTTTAAAAGTGCATTTAGGTATATAAACCAATCCAAATGCTG
TTGTTTGCTTTGCAGATCCCAAAGTGTCAGACAAGGAATTTCCAGGGTTGCATCCGCTAAGGTACAAGG
ACCTGCCAACTTCAGCATTTGGGCCATTAGAGAGTATACTCAAGGTTTACAGTGAGACTGTCAACATTC
GAACAGCTTCGGCAGTTATCATCAACTCAACAAGCTGTCTAGAGAGCTCATCTTTGGCATGGTTACAAA
AACAACTGCAAGTTCCAGTGTATCCTATAGGCCCACTTCACATTGCAGCTTCAGCGCCTTCTAGTTTAC
TTGAAGAGGACAGGAGTTGCCTTGAGTGGTTGAACAAGCAAAAAATAGGCTCAGTGATTTACATAAGTT
TGGGAAGCTTGGCTCTAATGGAAACTAAAGACATGTTGGAGATGGCTTGGGGTTTACGTAATAGCAACC
AACCTTTCTTATGGGTGATCCGACCGGGTTCTATTCCCGGCTCGGAATGGACAGAGTCTTTACCGGAGG
AATTCAGTAGGTTGGTTTCAGAAAGAGGTTACATTGTGAAATGGGCACCACAGATAGAAGTTCTCAGAC
ATCCTGCAGTGGGAGGGTTTTGGAGTCACTGCGGATGGAACTCGACCCTAGAGAGCATCGGGGAAGGAG
TTCCGATGATCTGTAGGCCTTTTACGGGAGATCAGAAAGTCAATGCGAGGTACTTAGAGAGAGTTTGGA
GAATTGGGGTTCAATTGGAAGGAGAGCTGGATAAAGGAACAGTGGAGAGAGCTGTAGAGAGATTGATTA
TGGATGAAGAAGGAGCAGAAATGAGGAAGAGAGTTATCAACTTGAAAGAGAAGCTTCAAGCCTCTGTCA
AGAGTAGAGGTTCCTCATTCAGCTCATTAGACAACTTTGTCAATTCCTTAAAAATGATGAATTTCATGT
AG

UGT76E11 Figure 27

ATGGAGGAAAAGCCGGCGGGCAGAAGAGTAGTGTTGGTTGCAGTTCCAGCTCAAGGACATATCTCTCCA
ATAATGCAACTTGCAAAAACACTTCACTTGAAGGGTTTCTCAATCACAATCGCTCAGACAAAGTTCAAT
TACTTTAGCCCTTCAGATGACTTCACTGATTTTCAGTTTGTCACCATTCCAGAAAGCTTACCAGAGTCT
GATTTTGAGGATCTCGGGCCAATAGAGTTTCTGCATAAGCTCAACAAAGAGTGTCAGGTGAGCTTCAAA
GACTGTTTGGGTCAGTTGTTGCTGCAACAAGGTAATGAGATAGCCTGTGTTGTCTACGACGAGTTCATG
TACTTTGCTGAAGCTGCAGCCAAAGAGTTTAAGCTTCCAAACGTCATTTTCAGCACCACAAGTGCCACG
GCTTTTGTTTGCCGCTCTGCATTCGACAAACTTTATGCAAACAGTATCCTGACTCCCTTGAAAGGTACT
CTTGAATTCTCTGTCTTCTATTCTTGCTGGTTTCTATAATCTGTAACAGCATGGTTCTTGACCTTTTTG
CAGAACCCAAAGGACAACAAAACGAGCTAGTGCCAGAGTTTCATCCCCTGAGATGCAAAGACTTTCCGG
TTTCACATTGGGCATCATTAGAAAGCATGATGGAGCTGTATAGGAATACAGTTGACAAACGGACAGCTT
CCTCGGTGATAATCAACACAGCGAGCTGTCTAGAGAGCTCATCTCTGTCTCGTCTGCAGCAACAGCTAC
AAATTCCAGTTTATCCTATAGGCCCTCTTCACCTGGTGGCATCAGCTTCTACGAGTCTTCTTGAAGAGA
ACAAGAGCTGTATTGAATGGTTGAACAAACAAAAGAAAAACTCTGTGATATTCGTAAGCTTGGGAAGCT
TAGCTTTGATGGAAATCAATGAGGTGATAGAAACTGCTTTGGGATTGGATAGTAGCAAGCAACAGTTCT
TGTGGGTCATTCGGCCAGGGTCAGTACGTGGTTCGGAATGGATAGAGAACTTGCCTAAGGAGTTTAGTA
AGATAATTTCGGGTCGAGGTTACATTGTGAAATGGGCTCCACAGAAGGAAGTACTTTCTCATCCTGCAG
TAGGAGGATTTTGGAGCCATTGCGGATGGAACTCGACACTAGAGAGCATCGGGGAAGGAGTTCCAATGA
TTTGCAAGCCGTTTTCCAGTGATCAAATGGTGAATGCGAGATACTTGGAGTGTGTATGGAAAATTGGGA
TTCAAGTTGAGGGTGATCTAGACAGAGGAGCGGTCGAGAGAGCTGTGAGGAGGTTAATGGTGGAGGAAG
AAGGGGAGGGGATGAGGAAGAGAGCTATCAGTTTGAAAGAGCAACTTAGAGCCTCTGTTATAAGTGGAG
GTTCTTCACACAACTCGCTAGAGGAGTTTGTACACTACATGAGGACTCTATGA

UGT76E12 Figure 28

ATGCAGGTTTTGGGAATGGAGGAAAAGCCTGCAAGGAGAAGCGTAGTGTTGGTTCCATTTCCAGCACAA
GGACATATATCTCCAATGATGCAACTTGCCAAAACCCTTCACTTAAAGGGTTTCTCGATCACAGTTGTT
CAGACTAAGTTCAATTACTTTAGCCCTTCAGATGACTTCACTCATGATTTTCAGTTCGTCACCATTCCA
GAAAGCTTACCAGAGTCTGATTTCAAGAATCTCGGACCAATACAGTTTCTGTTTAAGCTCAACAAAGAG
TGTAAGGTGAGCTTCAAGGACTGTTTGGGTCAGTTGGTGCTGCAACAAAGTAATGAGATCTCATGTGTC
ATCTACGATGAGTTCATGTACTTTGCTGAAGCTGCAGCCAAAGAGTGTAAGCTTCCAAACATCATTTTC
AGCACAACAAGTGCCACGGCTTTCGCTTGCCGCTCTGTATTTGACAAACTATATGCAAACAATGTCCAA
GCTCCCTTGAAAGGTACTCTAAAACTCTCTGTTTCGTGGTTTCCGCGAGTGGCTATAAGATTGAAACAG
CATTGTTTTTGACCTTTTTTGCAGAAACTAAAGGACAACAAGAAGAGCTAGTTCCGGAGTTTTATCCCT
TGAGATATAAAGACTTTCCAGTTTCACGGTTTGCATCATTAGAGAGCATAATGGAGGTGTATAGGAATA
CAGTTGACAAACGGACAGCTTCCTCGGTGATAATCAACACTGCGAGCTGTCTAGAGAGCTCATCTCTGT
CTTTTCTGCAACAACAACAGCTACAAATTCCAGTGTATCCTATAGGCCCTCTTCACATGGTGGCCTCAG
CTCCTACAAGTCTGCTTGAAGAGAACAAGAGCTGCATCGAATGGTTGAACAAACAAAAGGTAAACTCGG
TGATATACATAAGCATGGGAAGCATAGCTTTAATGGAAATCAACGAGATAATGGAAGTCGCGTCAGGAT
TGGCTGCTAGCAACCAACACTTCTTATGGGTGATCCGACCAGGGTCAATACCTGGTTCCGAGTGGATAG
AGTCCATGCCTGAAGAGTTTAGTAAGATGGTTTTGGACCGAGGTTACATTGTGAAATGGGCTCCACAGA
AGGAAGTACTTTCTCATCCTGCAGTAGGAGGGTTTTGGAGCCATTGTGGATGGAACTCGACACTAGAAA
GCATCGGCCAAGGAGTTCCAATGATCTGCAGGCCATTTTCGGGTGATCAAAAGGTGAACGCTAGATACT
TGGAGTGTGTATGGAAAATTGGGATTCAAGTGGAGGGTGAGCTAGACAGAGGAGTGGTCGAGAGAGCTG
TGAAGAGGTTAATGGTTGACGAAGAAGGAGAGGAGATGAGGAAGAGAGCTTTCAGTTTAAAAGAGCAAC
TTAGAGCCTCTGTTAAAAGTGGAGGCTCTTCACACAACTCGCTAGAAGAGTTTGTACACTTCATAAGGA
CTCTATGA

UGT76E2 Figure 29

ATGGAGGAAAAGCAAGTGAAGGAGACAAGGATAGTGTTGGTTCCAGTTCCAGCTCAAGGTCATGTAACT
CCGATGATGCAACTAGGAAAAGCTCTTCACTCAAAGGGTTTCTCCATCACTGTTGTTCTGACACAGTCT
AATCGAGTTAGCTCTTCCAAAGACTTCTCTGATTTCCATTTCCTCACCATCCCAGGCAGCTTAACTGAG
TCTGATCTCCAAAACCTAGGACCACAAAAGTTTGTGCTCAAGCTCAATCAAATTTGTGAGGCAAGCTTC
AAGCAGTGTATAGGTCAACTATTGCATGAACAATGTAATAATGATATTGCTTGTGTCGTCTACGATGAG
TACATGTACTTCTCTCATGCTGCAGTAAAAGAGTTTCAACTTCCTAGTGTCGTCTTTAGCACGACAAGT
GCTACTGCTTTTGTCTGTCGCTCTGTTTTGTCTAGAGTCAACGCAGAGTCGTTCTTGATCGACATGAAA
GGTATTCAAGATTCTAGCTTGTTTTATCTTAATTCAAAATCCTATTTATAGAAACTAATCCAAATGATC
GATGTTATCTTTTCAGATCCTGAAACACAAGACAAAGTATTTCCAGGGTTGCATCCTCTGAGGTACAAG
GATCTACCAACTTCAGTATTTGGGCCAATAGAGAGTACGCTCAAGGTTTACAGTGAGACTGTGAACACT
CGAACAGCTTCCGCTGTTATCATCAACTCAGCAAGCTGTTTAGAGAGCTCATCTTTGGCAAGGTTGCAA
CAACAACTGCAAGTTCCGGTGTATCCTATAGGCCCACTTCATATTACAGCTTCAGCGCCTTCTAGTTTA
CTAGAAGAAGACAGGAGTTGCGTTGAGTGGTTGAACAAGCAAAAATCAAATTCAGTTATTTACATAAGC
TTGGGAAGCTTGGCTCTAATGGACACCAAAGACATGTTGGAGATGGCTTGGGGATTAAGTAATAGCAAC
CAACCTTTCTTATGGGTGGTCAGACCGGGCTCTATTCCGGGGTCAGAATGGACAGAGTCCTTACCAGAG
GAATTCAATAGGTTGGTTTCAGAAAGAGGTTACATTGTGAAATGGGCTCCGCAGATGGAAGTTCTCAGA
CATCCTGCAGTAGGAGGGTTTTGGAGTCACTGTGGATGGAACTCAACAGTAGAGAGCATCGGGGAAGGA
GTTCCGATGATATGTAGGCCTTTCACCGGGGATCAGAAAGTCAATGCGAGGTACTTAGAGAGAGTTTGG
AGAATTGGGGTTCAATTGGAGGGAGATCTGGATAAAGAAACTGTGGAGAGAGCTGTAGAGTGGTTGCTT
GTGGATGAAGAAGGAGCAGAAATGAGGAAGAGAGCCATTGACTTGAAAGAAAAGATTGAAACCTCTGTT
AGAAGTGGAGGTTCCTCATGCAGCTCACTAGACGACTTTGTTAATTCCATGTGA

UGT78D1 Figure 30

ATGACCAAATTCTCCGAGCCAATCAGAGACTCCCACGTGGCAGTTCTCGCGTTTTTCCCCGTTGGCGCT
CATGCCGGTCCTCTCTTAGCCGTCACTCGCCGTCTCGCCGCCGCTTCTCCCTCCACCATCTTTTCTTTC
TTCAACACCGCAAGATCAAACGCGTCGTTGTTCTCCTCTGATCATCCCGAGAACATCAAGGTCCACGAC
GTCTCTGACGGTGTTCCGGAGGGAACCATGCTCGGGAATCCACTGGAGATGGTCGAGCTGTTTCTCGAA
GCGGCTCCACGTATTTTCCGGAGCGAAATCGCGGCGGCAGAGATAGAAGTTGGAAAGAAAGTGACATGC
ATGCTAACAGATGCCTTCTTCTGGTTCGCAGCGGACATAGCGGCTGAGCTGAACGCGACTTGGGTTGCC
TTCTGGGCCGGCGGAGCAAACTCACTCTGTGCTCATCTCTACACTGATCTCATCAGAGAAACCATCGGT
CTCAAAGGTAACTAGCTTTTTAGCGTTTAGTGATTATTCCACAAATTCAGCTACTACACTTTGTATGTA
TTTATGGTTATTATTATTATTTATCTCCTGGTAGATGTGAGTATGGAAGAGACATTAGGGTTTATACCA
GGAATGGAGAATTACAGAGTTAAAGATATACCAGAGGAAGTTGTATTTGAAGATTTGGACTCTGTTTTC
CCAAAGGCTTTATACCAAATGAGTCTTGCTTTACCTCGTGCCTCTGCTGTTTTCATCAGTTCCTTTGAA
GAGTTAGAACCTACATTGAACTATAACCTAAGATCCAAACTTAAACGTTTCTTGAACATCGCCCCTCTC
ACGTTATTATCTTCTACATCGGAGAAAGAGATGCGTGATCCTCATGGCTGCTTTGCTTGGATGGGGAAG
AGATCAGCTGCTTCTGTAGCGTACATTAGCTTCGGCACCGTCATGGAACCTCCTCCTGAAGAGCTTGTG
GCGATAGCACAAGGGTTGGAATCAAGCAAAGTGCCGTTTGTTTGGTCGCTGAAGGAGAAGAACATGGTT
CATCTACCAAAAGGGTTTTTGGATCGGACAAGAGAGCAAGGGATAGTGGTTCCTTGGGCTCCACAAGTG
GAACTGCTGAAACACGAGGCAATGGGTGTGAATGTGACACATTGTGGATGGAACTCAGTGTTGGAGAGT
GTGTCGGCAGGTGTACCGATGATCGGCAGACCGATTTTGGCGGATAATAGGCTCAACGGAAGAGCAGTG
GAGGTTGTGTGGAAGGTTGGAGTGATGATGGATAATGGAGTCTTCACGAAAGAAGGATTTGAGAAGTGT
TTGAATGATGTTTTTGTTCATGATGATGGTAAGACGATGAAGGCTAATGCCAAGAAGCTTAAAGAAAAA
CTCCAAGAAGATTTCTCCATGAAAGGAAGCTCTTTAGAGAATTTCAAAATATTGTTGGACGAAATTGTG
AAAGTTTAG

UGT89B1 Figure 31

ATGAAAGTGAACGAGGAAAACAACAAGCCGACAAAGACCCATGTCTTAATCTTCCCATTTCCGGCGCAA
GGTCACATGATTCCCCTCCTCGACTTCACCCACCGCCTTGCTCTCCGCGGCGGCGCCGCCTTAAAAATA
ACCGTCCTAGTCACTCCAAAAAACCTTCCTTTTCTCTCTCCGCTTCTCTCCGCCGTAGTTAACATCGAA
CCACTTATCCTCCCTTTTCCCTCCCACCCTTCAATCCCCTCCGGCGTCGAAAACGTCCAAGACTTACCT
CCTTCAGGCTTCCCTTTAATGATCCACGCGCTTGGTAATCTCCACGCGCCGCTTATCTCTTGGATTACT
TCTCACCCTTCTCCTCCAGTAGCCATCGTATCTGATTTCTTCCTTGGTTGGACCAAAAACCTCGGAATC
CCTCGTTTCGATTTCTCTCCCTCCGCTGCTATCACTTGCTGCATACTCAATACTCTCTGGATCGAAATG
CCCACCAAGATCAACGAAGATGACGATAACGAGATCCTCCACTTTCCCAAGATCCCGAATTGTCCAAAA
TACCGTTTTGATCAGATCTCCTCTCTTTACAGAAGTTACGTTCACGGAGATCCAGCTTGGGAGTTCATA
AGAGACTCCTTTAGAGATAACGTGGCGAGTTGGGGACTCGTCGTGAACTCGTTCACCGCCATGGAAGGT
GTTTATCTCGAACATCTTAAGCGAGAGATGGGCCATGATCGTGTATGGGCTGTAGGCCCAATTATTCCG
TTATCTGGGGATAACCGTGGTGGCCCGACTTCTGTTTCTGTTGATCACGTGATGTCGTGGCTTGACGCA
CGTGAGGATAACCACGTGGTGTACGTGTGCTTTGGAAGTCAAGTAGTTTTGACTAAAGAGCAGACTCTT
GCACTCGCCTCTGGGCTTGAGAAAAGCGGCGTCCATTTCATATGGGCCGTAAAGGAGCCCGTTGAGAAA
GACTCAACACGTGGCAACATCCTGGACGGTTTCGACGATCGCGTGGCTGGGAGAGGTCTGGTGATCAGA
GGATGGGCTCCACAAGTAGCTGTGCTACGTCACCGAGCCGTTGGCGCGTTTTTAACGCACTGTGGTTGG
AACTCTGTGGTGGAGGCGGTTGTCGCCGGCGTTTTGATGCTGACGTGGCCGATGAGAGCTGACCAGTAC
ACTGACGCGTCTCTGGTGGTTGATGAGTTGAAAGTAGGTGTGCGTGCTTGCGAAGGACCTGACACGGTG
CCTGACCCGGACGAGTTAGCTCGAGTTTTCGCTGATTCCGTGACCGGAAATCAAACGGAGAGGATCAAA
GCCGTGGAGCTGAGGAAAGCAGCGTTGGATGCGATTCAAGAACGTGGGAGCTCAGTGAATGATTTAGAT
GGATTTATCCAACATGTCGTTAGTTTAGGACTAAACAAATGA

72B3 ORF Figure 32

ATGAGCATAGATATTTTTCAAGAAATAAGAATAAAGAAAATTCTACTCTTAATGGCGGAAGCAAACACT
CCACACATAGCAATCATGCCGAGTCCCGGTATGGGTCACCTTATCCCATTCGTCGAGTTAGCAAAGCGA
CTCGTTCAGCACGACTGTTTCACCGTCACAATGATCATCTCCGGTGAAACTTCGCCGTCTAAGGCACAA
AGATCCGTTCTCAACTCTCTCCCTTCCTCCATAGCCTCCGTATTTCTCCCTCCCGCCGATCTTTCCGAT
GTTCCCTCCACAGCGCGAATCGAAACTCGGGCCATGCTCACCATGACTCGTTCCAATCCGGCGCTCCGG
GAGCTTTTTGGCTCTTTATCAACGAAGAAAAGTCTCCCGGCGGTTCTCGTCGTCGATATGTTTGGTGCG
GATGCGTTCGACGTGGCCGTTGACTTCCACGTGTCACCATACATTTTCTATGCATCCAATGCAAACGTC
TTGTCGTTTTTTCTTCACTTGCCGAAACTAGACAAAACGGTGTCGTGTGAGTTTAGGTACTTAACCGAA
CCGCTTAAGATTCCCGGCTGTGTCCCGATAACCGGTAAGGACTTTCTTGATACGGTTCAAGACCGAAAC
GACGACGCATACAAATTGCTTCTCCATAACACCAAGAGGTACAAAGAAGCTAAAGGGATTCTAGTGAAT
TCCTTCGTTGATTTAGAGTCGAATGCAATAAAGGCCTTACAAGAACCGGCTCCTGATAAACCAACGGTA
TACCCGATTGGGCCGCTGGTTAACACAAGTTCATCTAATGTTAACTTGGAAGACAAGTTCGGATGTTTA
AGTTGGCTAGACAACCAACCATTCGGCTCGGTTCTATACATATCATTTGGAAGCGGCGGAACACTTACA
TGTGAGCAGTTTAATGAGCTTGCTATTGGTCTTGCGGAGAGCGGAAAACGGTTTATTTGGGTCATACGA
AGTCCAAGCGAGATAGTTAGTTCGTCGTATTTCAATCCACACAGCGAGACAGACCCCTTTTCGTTTTTA
CCAATTGGGTTCTTAGACCGAACCAAAGAGAAAGGTTTGGTGGTTCCATCATGGGCTCCACAGGTTCAA
ATCCTGGCTCATCCATCCACATGCGGGTTTTTAACACACTGTGGATGGAATTCGACCTTAGAAAGCATT
GTAAACGGTGTACCACTCATAGCGTGGCCTTTATTCGCGGAGCAAAAGATGAATACATTGCTACTCGTG
GAGGATGTTGGAGCGGCTCTAAGAATCCATGCGGGTGAAGATGGGATTGTACGGAGGGAAGAAGTGGTG
AGAGTGGTGAAGGCACTGATGGAAGGTGAAGAGGGAAAAGCCATAGGAAATAAAGTGAAGGAGTTGAAA
GAAGGAGTTGTTAGAGTCTTGGGTGACGATGGATTGTCCAGCAAGTCATTTGGTGAAGTTTTGTTAAAG
TGGAAAACGCACCAGCGAGATATCAACCAAGAGACGTCCCACTAA

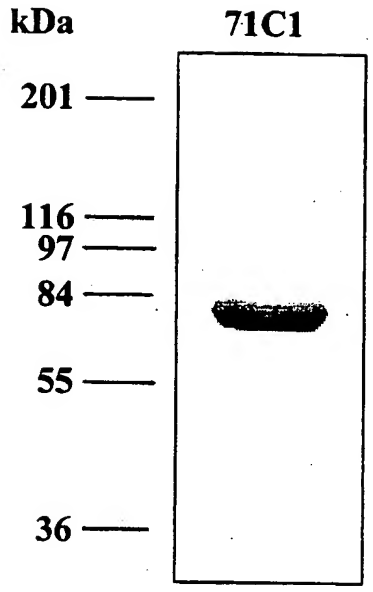

Figure 33

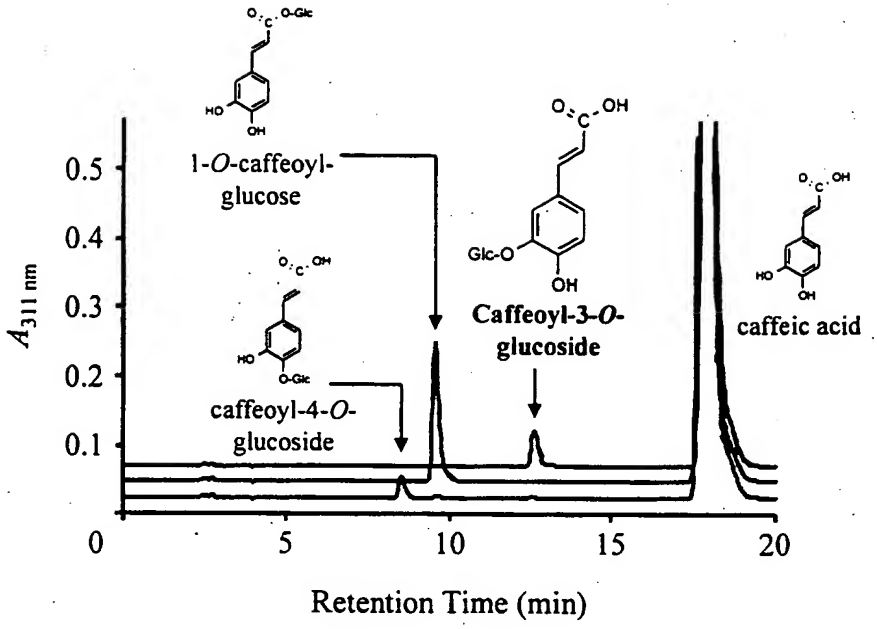
1-O-caffeoyl-glucose
caffeoyl-4-O-glucoside
Caffeoyl-3-O-glucoside
caffeic acid
A311 nm
0.5
0.4
0.3
0.2
0.1
0
5
10
15
20
Retention Time (min)

Figure 34

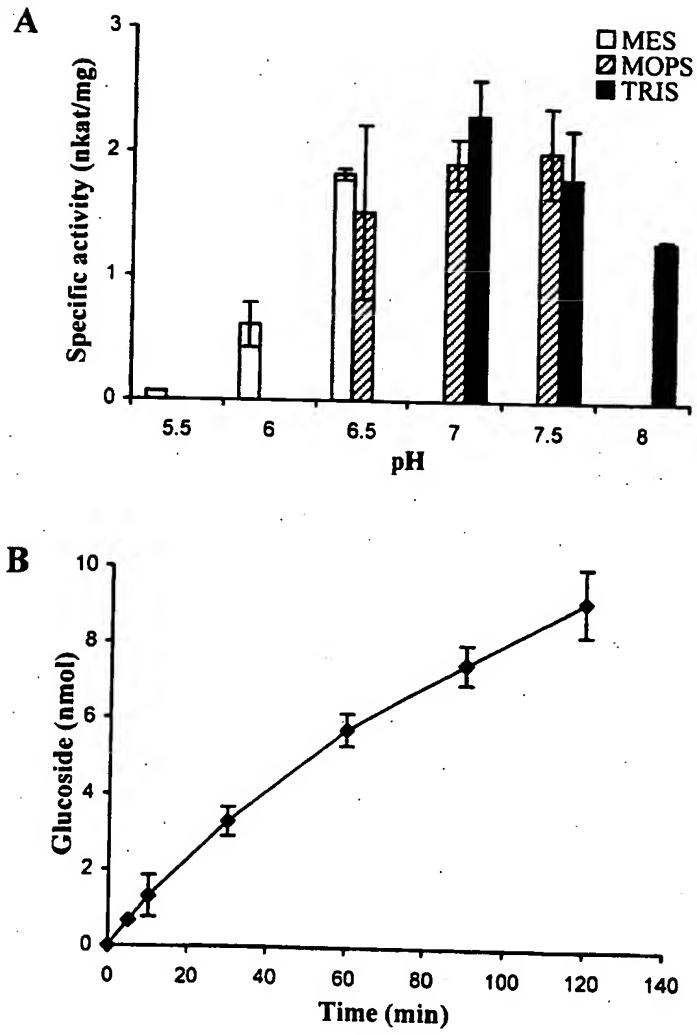
A
Specific activity (nkat/mg)
MES
MOPS
TRIS
pH
B
Glucoside (nmol)
Time (min)

Figure 35

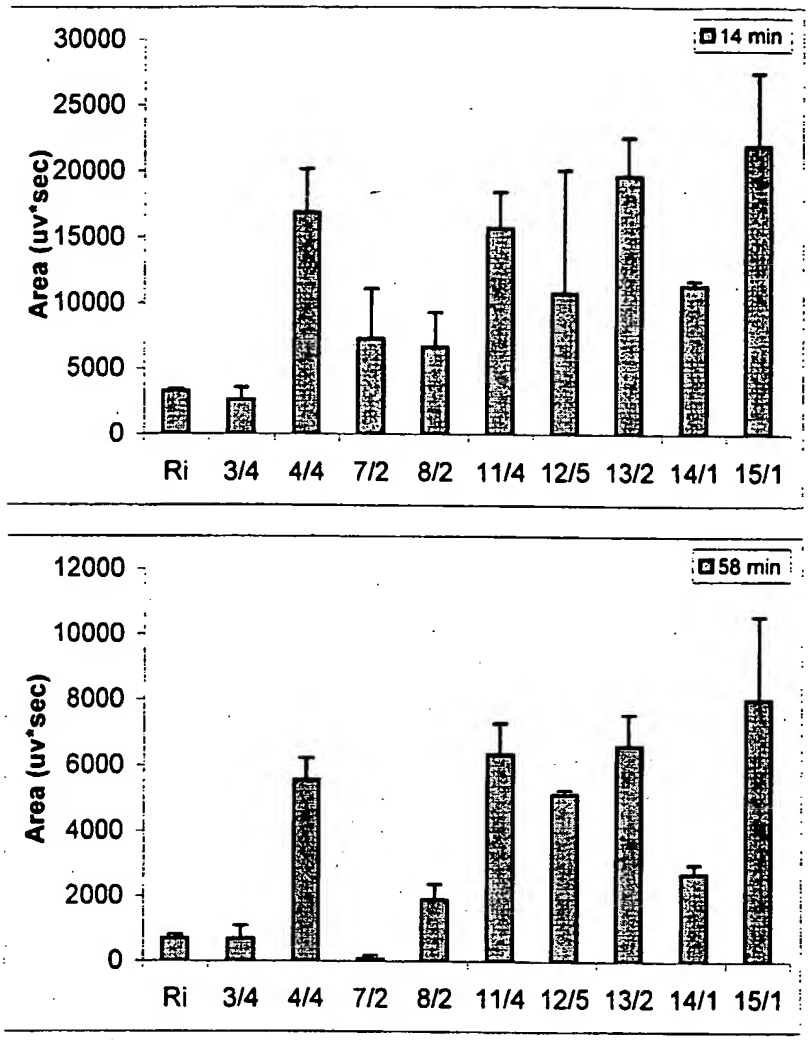
14 min
Area (uv*sec)
30000
25000
20000
15000
10000
5000
0
Ri
3/4
4/4
7/2
8/2
11/4
12/5
13/2
14/1
15/1
58 min
Area (uv*sec)
12000
10000
8000
6000
4000
2000
0
Ri
3/4
4/4
7/2
8/2
11/4
12/5
13/2
14/1
15/1

Figure 36

UGT	Figure	LUTEOLIN	QUERCETIN	CATECHIN	CYANIDIN	CAFFEIC ACID		
71B5	12		+			-		
71C1	10A	+	++			+		
71C2	11A	+				-		
71C3	13		+					
71C5	14		+					
71D1	15		++			-		
72B3	32		+					
72E2	7A					+		
72E3	9A					+		
73B1	16	+	+			-		
73B2	17	+	+			-		
73B3	8A	+	++			+		
73B4	18	+	++					
73B5	19		++					
73C1	20	+				-		
73C3	21	+				-		
73C5	22	+	+			-		
73C6	23	+	+			-		
73C7	24	+						
74F2	25	+				-		
75B1	1A					+		
76E1	26	+						
76E11	27	+	+					
76E12	28	+	++					
76E2	29	+						
78D1	30		+					
84A1	4A					+		
84A4	3A					+		
84B1	5A	+	+			+		
89B1	31	+++	++					
+++	= highest activity							
++	= more than 20% activity							
+	= less than 20% activity							

Figure 37

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/GB 01/00477

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/82 C12N9/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
BIOSIS, EPO-Internal, WPI Data, PAJ, STRAND

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 967 283 A (SUNTORY LTD) 29 December 1999 (1999-12-29) the whole document	1,2,4,6, 10,14-16
X	WO 97 21816 A (ZENECA LTD ;MANNING KENNETH (GB)) 19 June 1997 (1997-06-19) the whole document	1,2,4,6, 10,14-16
X	WO 97 16559 A (PLANT GENETIC SYSTEMS NV ;CANADA NAT RES COUNCIL (CA); AUDENHOVE K) 9 May 1997 (1997-05-09) the whole document	1,2,4,6, 10,14-16
X	WO 00 00626 A (CHONG JULIE ;BALTZ RACHEL (FR); BEFFA ROLAND (FR); FRITIG BERNARD) 6 January 2000 (2000-01-06) the whole document	1,2,4,6, 10,14-16

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 July 2001	27/07/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Holtorf, S

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DATABASE EMBL SEQUENCE LIBRARY 'Online! 18 June 1998 (1998-06-18) JOHNSON-HOPSON, C., ET AL. : "genomic sequence for Arabidopsis thaliana BAC T25N20 from chromosome I - unpublished" XP002172235 accession no. AC005106 ---	
A	DATABASE EMBL SEQUENCE LIBRARY 'Online! 16 November 1998 (1998-11-16) NAKAMURA, Y.: "structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 Bp covered by sixty P1 and TAC clones" XP002172236 accession no. AB019232 ---	
A	DATABASE EMBL SEQUENCE LIBRARY 'Online! 4 July 1997 (1997-07-04) BEVAN, M., ET AL. : "EU Arabidopsis sequencing project - chromosome 4" XP002172237 Z97339 ---	
A	DATABASE EMBL SEQUENCE LIBRARY 'Online! 23 February 1998 (1998-02-23) BEVAN, M., ETAL. : "EU Arabidopsis Sequencing project" XP002172238 accession no. AL021961 ---	
A	DATABASE EMBL SEQUENCE LIBRARY 'Online! 7 October 1998 (1998-10-07) NAKAMURA, Y. : "Arabidopsis thaliana genomic DNA, chromosome 5" XP002172244 accession no. AB018119 ---	
A	DATABASE EMBL SEQUENCE LIBRARY 'Online! 19 August 1998 (1998-08-19) LIN,X., ET AL. : "Sequence analysis of chromosoem 2 of the plant Arabidopsis thaliana" XP002172239 accession no. AC005496 ---	
A	DATABASE EMBL SEQUENCE LIBRARY 'Online! 8 August 1997 (1997-08-08) LIN, X., ET AL. : "sequence analysis of chromosome 2 of the plant Arabidopsis thaliana" XP002172240 accession no. AC002391 ---	

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	GRAHAM RICHARD ET AL: "DNA sequence of UDP glucose indole-3-acetate beta-D-glucosyltransferase from Arabidopsis thaliana (Accession No. U81293)." PLANT PHYSIOLOGY (ROCKVILLE), vol. 113, no. 3, 1997, page 1004 XP002172231 ISSN: 0032-0889 ---	
T	JACKSON ROSAMOND G ET AL: "Identification and biochemical characterization of an Arabidopsis indole-3-acetic acid glucosyltransferase." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 276, no. 6, 9 February 2001 (2001-02-09), pages 4350-4356, XP002172232 ISSN: 0021-9258 ---	
T	LIM ENG-KIAT ET AL: "Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in Arabidopsis." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 276, no. 6, 9 February 2001 (2001-02-09), pages 4344-4349, XP002172233 ISSN: 0021-9258 ---	
T	LI YI ET AL: "Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 276, no. 6, 9 February 2001 (2001-02-09), pages 4338-4343, XP002172234 ISSN: 0021-9258 -----	

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0967283 A	29-12-1999	AU 8243298 A	16-02-1999
		CN 1237206 T	01-12-1999
		WO 9905287 A	04-02-1999
WO 9721816 A	19-06-1997	AU 1106297 A	03-07-1997
WO 9716559 A	09-05-1997	EP 0771878 A	07-05-1997
		AU 727492 B	14-12-2000
		AU 7496696 A	22-05-1997
		EP 0858511 A	19-08-1998
		JP 11514880 T	21-12-1999
WO 0000626 A	06-01-2000	FR 2780415 A	31-12-1999
		AU 4375999 A	17-01-2000